



THE BRAIN'S REPRESENTATIONAL POWER

ON CONSCIOUSNESS
AND THE INTEGRATION OF MODALITIES

CYRIEL M. A. PENNARTZ

The Brain's Representational Power

The Brain's Representational Power

On Consciousness and the Integration of Modalities

Cyriel M. A. Pennartz

**The MIT Press
Cambridge, Massachusetts
London, England**

© 2015 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu

This book was set in Syntax LT and Times New Roman by Toppan Best-set Premedia Limited. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Pennartz, Cyriel M. A., 1963—

The brain's representational power : on consciousness and the integration of modalities / Cyriel M.A. Pennartz.
pages cm

Includes bibliographical references and index.

ISBN 978-0-262-02931-5 (hardcover : alk. paper) 1. Neural circuitry. 2. Neurosciences. 3. Cognition.

4. Cognitive science. 5. Memory. I. Title.

QP363.3.P46 2015

612.8—dc23

2015001907

10 9 8 7 6 5 4 3 2 1

For Hanneke, Charlotte, and Flora

Contents

Preface and Acknowledgments	ix
1 Introduction	1
2 Building Representations with Neurons and Spikes	11
3 Peeling Off Our Conscious Lives	47
4 What Neural Network Models Can and Cannot Explain about Cognition	83
5 Networks and the Problem of Panpsychism	109
6 Structure and Function of Brain Systems for Conscious and Nonconscious Representation	131
7 Same Brain, Different States: Waking, Sleeping, and Anesthesia	173
8 Requirements for Conscious Representations	209
9 Neural Mechanisms for Conscious Representations	233
10 Levels of Representational Organization	265
11 Philosophical and Future Perspectives	285
Notes	317
References	323
Index	371

Preface and Acknowledgments

Having finished this book naturally comes with feelings of joy, but I am also struck by a paradox, in that the subjects treated here may be considered somewhat unrelated to most of the work I have done so far, which has skipped from slice physiology to neural network models, synaptic plasticity, circadian clocks, and to mechanisms of memory, sleep, and motivation as studied in animals. Only in recent years have I started experimental research on the relationships between neural coding, perception, representation, and multimodal integration. At the same time, it is also entirely logical that this book got there, not only as a theoretical exploration of uncharted territories that demand further empirical ventures, but also for two other reasons. The first originated from the functional nature precisely of those brain structures we tend *not* to associate with higher-order cognitive processes, culminating in that pinnacle we call consciousness. What makes brain structures such as the hypothalamus and basal ganglia so different from corticothalamic systems that only the latter are granted the privilege of staging conscious processing? Having long experienced this pleasant “itch” arising when two contrasting fields rub against one another, it is hardly surprising that I felt compelled to pursue the matter and was driven toward review, synthesis, and writing. Over the years this practice became more comprehensive, ramifying into questions which I felt were neither fully addressed in neuroscience (for being too philosophical) nor in philosophy (for being too neuroscientific). Not entirely without reason, some philosophers regard neuroscientists as pure behaviorists, and it seemed timely to counterbalance this view by exploring arguments and evidence for representational and world-modeling systems in the brain. A second example pertains to some principles of sensory coding which appear to be happily adopted in some major textbooks on neuroscience but give rise to deep philosophical problems on closer scrutiny, touching the philosopher’s raw nerve referred to as “perceptual qualia.” Likewise it seemed appropriate to recalibrate the role of computational neuroscience along a scale ranging from being utterly indispensable to being limited in several respects.

The second reason stems from personal history. When mind-traveling backward in time and trying to reconstruct where and how some of the ideas in this book arose, a first way station lies at the California Institute of Technology, where I was a postdoctoral

fellow around 1993. The confluence of ideas from computational physics, psychology, and experimental neurobiology helped shape concepts presented here, and these have been simmering and ripening over the past 20 years. Traveling further back, I recall having done a first undergraduate internship in neurobiology with Fernando Lopes da Silva at the University of Amsterdam around 1987 and hesitating between a neurochemical research project and a second bachelor's program in philosophy. The combination of neurobiology and philosophy raised some eyebrows at the time, but for me it proved to be the better, challenging choice and I thank Fernando for his understanding. Further back through the fog of time a boy emerges, making crayon drawings in an attempt to grasp the workings of the body organs, which were fancifully explained by all sorts of homunculi taking care of food digestion, respiration, muscular movement, and—why not?—thinking. This solution gave me a slight sense of unease that something in this account was not quite right, and particularly the problem of thinking-by-gray-matter left behind a tickling background feeling which did not fade away over time but tended to return with surprising power and force. Coming back to the present, the sense of paradox dissolves into the logic of seeing the branches from one's own history come together at the crossroads of the present book.

This book would not have become what it is today without the stimulating influences and general intellectual inputs provided by a great many colleagues over the years: Carol Barnes, Peter Boeijinga, Gerard Borst, Carlos Brody, Emery Brown, György Buzsaki, Matteo Carandini, Paul Churchland, Guy Debrock, Andreas Engel, Barry Everitt, Walter Freeman, Pascal Fries, Karl Friston, Ann Graybiel, Wim van de Grind, Henk Groenewegen, Ken Harris, Bruce Hope, John Hopfield, Rutsuko Ito, Ole Jensen, Marcel de Jeu, Marian Joëls, Christof Koch, Peter König, Victor Lamme, John Lisman, Ton Lohman, Fernando Lopes da Silva, Bruce McNaughton, Gloria Meredith, Earl Miller, Richard Morris, Edvard and May-Britt Moser, Tonny Mulder, John O'Keefe, Tjeerd olde Scheper, Jaap van Pelt, Giovanni Pezzulo, Jan Pieter Pijn, Rodrigo Quiroga, David Redish, Trevor Robbins, Barry Roberts, Pieter Roelfsema, Martin Sarter, Geoff Schoenbaum, Anil Seth, Wolf Singer, Henk Spekreijse, Stefano Taverna, Alessandro Treves, Harry Uylings, Paul Verschure, Pieter Voorn, Richard van Wezel, Menno Witter, Chris de Zeeuw, and Semir Zeki. Errors or weaknesses in this piece of work should by no means be attributed to them. I would also like to thank many of their coworkers for discussions. Much of contemporary neuroscience appears to gravitate, explicitly or implicitly, toward a nucleus of reductive physicalism and functionalism, and this work may be viewed as sailing against the prevailing wind. For this reason I should emphasize that by these acknowledgments no sympathy whatsoever is implied for the opinions expressed here. As the work both broadened and deepened while it was being written, it also became harder to provide full credits to all researchers who have contributed to the vast literature, ranging from neurobiology to philosophy, and so I should apologize for not referencing a number of studies which may have been relevant.

Over the past 10 years I have had the great privilege of working with a wonderful, multidisciplinary team of talented, enthusiastic, and assertive people who created a sociable and scintillating climate at our lab at the Science Faculty of the University of Amsterdam. Of the many coworkers who worked, or are still working, at the lab I would like to thank Tara Arbab, Francesco Battaglia, Jeroen Bos, Conrado Bosman-Vittini, Henrique Cabral, Sander Daselaar, Yvette van Dongen, Esther van Duuren, Luc Gentet, Wim Ghijssen, Pieter Goltstein, Willem Huijbers, Jadin Jackson, Ruud Joosten, Tobias Kalenscher, Lianne Klaver, Gerben Klein, Jan Lankelma, Carien Lansink, Eunjeong Lee, Jeannette Lorteije, Hemi Malkki, Matthijs van der Meer, Guido Meijer, Paul Mertens, Ivana Milojevic, Jorrit Montijn, Laura van Mourik-Donga, Rebecca Nordquist, Umberto Olcese, Quentin Perrenoud, Silviu Rusu, Martin Vinck, and Marijn van Wingerden. The efforts of my colleagues Paul Lucassen, Marten Smidt, and Wytse Wadman in sustaining an atmosphere conducive to science and education (while minimizing bureaucracy) are greatly appreciated. Thanks also to Willem Stiekema and Karen Tensen for helping to free up sabbatical time to work on this book, and to Aly Eekhof and Casper Huijser for making life easier in this respect. I am deeply grateful to Bob Prior, Chris Eyer, and Katherine Almeida of MIT Press for embarking on this publishing journey together and reminding me to stay on course when distracted by scientific sirens and other matters of ephemeral relevance along the way.

Students taking classes in cognitive neuroscience, systems neuroscience, and network modeling at the University of Amsterdam have already brought up many interesting questions and intriguing views on issues expressed here. I would like to encourage them, as well as students elsewhere, to keep on wondering, keep on asking, keep on pursuing mind–brain problems which—despite their venerable age—are becoming more empirically tractable in recent times. From the inexhaustible source of questions, try to tease out which ones are scientifically answerable and which are not. Keep your eyes and ears open, stay broadly tuned, and disobey the classical boundaries between disciplines. Although the vast landscape of science prevents us from being a true *homo universalis*, we can aspire to be one.

Once more wandering back in time, I feel very fortunate having grown up with parents, Paul and Ansje, who have been and still are encouraging, stimulating musings on the human body and much more. Especially since my father passed away, his own work on and interest in philosophy and the social sciences have gained deeper meaning, also in relation to this book. Finally, I am deeply grateful to my love, Hanneke, and my two daughters, Charlotte and Flora, for enabling me to work on this book with vigor and passion, sometimes bordering on the obsessive. A special thanks to Charlotte for her great sculpting of many illustrations. This book cannot make up for loss of the time we could have otherwise spent together.

After reading the pages to follow I hope you will share, even more than before, the excitement about being a hunter-gatherer—not of food, but of information—chasing the mysteries of human existence.

1

Introduction

I, an American writer, writing about America, was working from memory, and the memory is at best a faulty, warpy reservoir. I had not heard the speech of America, smelled the grass and trees and sewage, seen its hills and water, its color and quality of light. I knew the changes only from books and newspapers. But more than this, I had not felt the country for twenty-five years. In short, I was writing of something I did not know about, and it seems to me that in a so-called writer this is criminal.

—John Steinbeck (1962)

1.1 Introduction

As compared to the philosophical inquiries on mind–brain relationships over the past centuries, research in psychology, experimental neuroscience, and computational modeling has accelerated at a breathtaking pace over the past few decades. This book deals with the neural basis of consciousness and, more specifically, with the foundations of neural representations underlying consciousness. In this branch of mind–brain research, much recent work has been spawned from psychology, particularly by the multidisciplinary field called “cognitive science,” but the input from neuroscience has remained somewhat underexposed so far. In this book I deliberately adopt a neuroscientific angle to the question of how neural systems generate representations and consciousness, but at the same time I attempt to pay attention to explaining the “hardest” aspect of consciousness. In contrast to aspects that have proven accessible to experimental study, such as attention, memory, decision making, and behavioral reporting about one’s experiences, this hardest aspect is considered by many to be the subjective, qualitative nature of conscious experience, and precisely because of this nature it has been vexingly difficult to come up with plausible neural explanations. The approach I chose for this book is, first, to isolate brain systems and physiological processes that can be selectively associated with consciousness. Based on neurological and other neuroscientific findings, this will next lead us to examine the neural basis of representations in the brain, particularly those in the sensory domain. I will argue that not just any type of neural activity pattern is sufficient to generate the sort of sensory representation we can associate with consciousness. This will prompt us to examine

the requirements that neural systems should fulfill to qualify as generators of the kind of experiences we are aware of. Finally, we will ask how those requirements may be neurally implemented. It will turn out to be important to parse the problem of neural representations at multiple levels of neural organization—at the level of single cells, groups of neurons, and even more complex aggregates.

What is consciousness, and how may the concept of “representation” be informative about it? Many of us have a clear intuition about when we are conscious ourselves, what we experience during our conscious lives, and when we think people around us are conscious or unconscious. Yet, even though we seem to have a keen sense of what consciousness is about, a clear definition eludes us. If we try to pin down consciousness by the fact that we live a life full of rich experiences, we stumble on the difficulty of defining “experience” without resorting to circularity. When we think of consciousness as a basic state of “knowing” about things around us or about ourselves, we are confronted with a gap in our understanding of how brains generate knowledge in the first place, and how even the most familiar and common objects in our environment become meaningful to us. Objects become meaningful in part because we learn to relate them to other objects and events that we encounter in our daily lives. Starting in early childhood, we learn to weave an object into the fabric of things we associate with that object, such as when we link the word “book” to reading, storytelling, characters, and paper. But we also run up against the question of how an object—no matter how new or strange—can have a direct content when we first perceive it. How is it possible that an object has this qualitative content that is manifested to us in all its brusqueness when we first perceive it? Every time we feel consciousness is within our reach, it makes a getaway, like an octopus escaping into the darkness of the sea, leaving us behind with the frustration of self-referential definition.

When we attempt to find out how consciousness is linked to brain function, it seems more useful to study how conscious experience manifests itself, how we apply our vocabulary to describe subjective experiences of others and ourselves, and how we link it to empirical observations on brains in all their physical detail—taking into account how our descriptions are tainted by linguistic habits. We will have to ask when we consider ourselves and our fellow citizens to be conscious, and when not. We will analyze what the bare necessities are to achieve a minimal amount of consciousness (if there is any such threshold) and which living beings or artifacts may be thought of as conscious. How far does one have to go down the scale of evolutionary complexity before it is no longer reasonable to consider a living creature conscious? Even common daily-life situations seduce us to make distinctions between conscious and nonconscious beings, for instance, in our choice of food and in our interior monologues while struggling with the idiosyncrasies of personal computers. However, if we wish to regard a chimpanzee as conscious but a shrimp or palmtop as nonconscious, it is of equal interest to know whether we are right or wrong as to learn about the criteria that lead us to make the distinction.

In addition to this evolutionary perspective, current neuroscience offers a wealth of neurological and experimental data to empirically address brain–consciousness relationships. When our brains shift from a state of consciousness into sleep or deep anesthesia, we can safely say that the brain’s anatomical structures and synaptic connections remain intact—so what is it in the neurophysiology of the brain that corresponds to this change in state? This question will bring us also to neurological evidence gathered in patients suffering from brain damage, highlighting which brain structures and brain functions are associated with consciousness. For instance, the cerebellum contains more than half of the human brain’s estimated total of more than 10^{11} neurons and thus vastly outnumbers the mass of neurons in the neocortex (or cerebral cortex), yet it is this cortex that is implicated in conscious processing. What makes the cerebral cortex so special?

The question of what functional and circuit properties a neural system should have to sustain consciousness can be equally well posed for artificial systems. Does IBM’s supercomputer Watson fulfill any basic requirements to qualify for conscious experience, or should we rather think of smartphones as leading the way toward an artificial conscious system? Which computer technology is most brain-like, and how could we define criteria to test whether an artifact produces even the slightest glimpse of consciousness?

A modern theory of mind–brain relationships is not complete without providing firm and dauntless answers to these questions. That being said, giving firm answers is not so much of a problem as validating them. The field of brain–mind research is so inundated with conceptual difficulties, philosophical controversy, and empirical unknowns that even the firmest answer should be regarded with caution. Personally, I found it extremely challenging and thrilling to attempt to put aside some of our most basic intuitions about consciousness, which we might secretly cherish as a kind of internal “holy grail” when facing the physical nature of the world around us. Valuing intuitions against empirical findings is certainly not entirely free of emotion, and here it is particularly hard—but necessary—to take Richard Feynman’s *adagium* of being disinterested in one’s own scientific results to heart. If we are inclined to say that a honeybee or an iPhone can never be conscious whereas the human brain clearly can, we have to scrutinize whether this attitude arises from latent emotions that may have to do with a grudge against being reduced to the dull world of a digital device—or whether there are good reasons to uphold the distinction. Even if we would succeed in being emotionally neutral about the outcome of a conscious–nonconscious dichotomy, thinking about the mind–brain inevitably entails making conscientious decisions about which criteria to employ to consider some systems conscious or not, and which type of evidence should be weighed heavily or lightly. All of this, however, should not diminish the intellectual joy in undertaking the journey.

1.2 Some Inevitable Distinctions—The Zoo of Language

Our language presents us with a perplexing zoo of expressions and underlying concepts that illustrate the various ways we think about brains and consciousness. Let us begin to make a quick inventory of terms and to seek at least an initial justification for them against a background of common, daily-life assumptions. When you see on television how a boxer is knocked unconscious during a fight, should consciousness be thought of as the same phenomenon as when you awake early in the morning and become aware¹ of the room you are lying in? In the first situation we use the term “consciousness” by virtue of external observation of a person who, to his misfortune, displays behavioral signs of losing consciousness: his body falls to the ground, his head hits the floor without restraint or control, his eyes close, and his body remains motionless. Considering the boxer “unconscious” is a defendable, rational use of the term because there is clear evidence for a particular mind–brain state, albeit evidence that is indirect, obtained by behavioral observation. A stuntman could imitate the boxer’s downfall with great precision without being knocked out. Because we cannot exchange our own brain with the boxer’s intracranial mass, it is impossible to experience directly what the boxer feels when being hit or when regaining awareness—later realizing that a lapse of consciousness occurred. Here we will accept the use of “consciousness” to describe the boxer’s mind–brain state as rational and reasonable, noting that the boxer’s unconscious state was inferred from his behavior, taking into account the great similarity between his body and behavior and our own.

If you wake up after a good night’s sleep and tell me that the first thing you became aware of is a painting on the wall depicting a lady with an umbrella in a field full of poppies, I have little reason to doubt you underwent a transition in conscious state. You describe a visual, phenomenal experience—which to me, as observer, is not expressed via body language as in the case of the boxer, but communicated by voice. Despite the credibility of your vivid utterance, there is a distinction to make in the kind of evidence that you and I have about your conscious state. You are having a direct experience of the lady in the poppy field which gives little reason to doubt it yourself. I will refer to this kind of experience, full of colors, shapes, textures—all framed in space and time and ordered in recognizable objects or background elements such as a lady and the poppy field—as a phenomenal experience, or simply an experience. Of course, one might ask whether this image was truly the first experiential content you became aware of when waking up—perhaps a vaguer gist of the surrounding room preceded it. Regardless of this, for yourself there may be no better way of gathering evidence for your conscious experience—even the best current techniques for measuring brain activity could not fill in for you what you were exactly experiencing when waking up.

In contrast, after hearing you talk, I can imagine what it would be like for you to wake up and see the painting, but in doing so, I am converting indirectly gathered evidence, communicated by you to me, into a substitute experience. Your voice is cueing my brain systems

to construct and imagine this substitute, which may not correspond to how you experienced the image yourself. For me, the evidence on your conscious experience is indirect, as in the boxer’s case, even though language allows us to describe our own experiences in richer detail and with greater vividness than could be evoked by whole-body movements. No matter how strong my belief in what you say, you *could* be lying to me, or your statement could be involuntarily inaccurate. Your mind–brain state could be different from what I would guess it is. I could turn around to look at the wall opposite your bed and see no painting hanging there at all—leading me to conclude you are suffering from a delusion. However, these are exceptional cases. My evidence on your experience may be indirect, but it is generally meaningful and valid to consider you being conscious. Your behavior, language, and body structure are utterly similar to mine, and this very strong analogy is a powerful instrument for interpreting your utterance as reflecting a conscious state.

I am avoiding the use of the term “introspection” here in delineating different kinds of evidence on consciousness. Classifying your description of the painting as “introspective” would beg the question what “extraspective” would mean. Your perception is as “extraspective” as would be your participation in experimental measurements where you have to count people passing through a street, record electrical activity from brain cells, or count collision events in a particle accelerator. “Introspection” is frequently used as a way to peek into the “subjectivity” of personal experiences. At first your statement on the painting may be considered subjective, but once I and others start verifying and confirming it, it gains in objectivity. The division between subjectivity and objectivity hinges on the reproducibility of our individual observations by others. Nonetheless, the subjectivity of individual experience is an aspect that neural explanations of consciousness will have to cope with.

Assuming most people will go along with the use of “consciousness” as describing someone’s mind–brain state based on various kinds of evidence, what would we classify as a nonsensical use? An assertion like “The beach is having a conscious recollection of my walk along it yesterday” does not make sense—except perhaps as poetry. “The tree was conscious of the wind as it was blowing through its branches and leaves, waving back and forth in unison.” Is it rational to apply the notion of consciousness here? I believe a scientific answer to this should be negative. It is unreasonable to consider the beach and the tree conscious. However, later on we will have to face the question why this is exactly the case. Stunningly, even modern theories of consciousness typically fail to explain why a brain, viewed as a densely interconnected network of cells, can be conscious whereas a tree or a beach will never be conscious. This refers to the problem of panpsychism, and I will return to it in various chapters. Panpsychism—a composite of the Greek words “pan” (all) and “psyche” (soul, mind)—expresses the belief that all objects in nature have mind-like and even conscious capacities. If a theory of consciousness implies that it is present virtually everywhere throughout nature, the problem occurs that the term “consciousness” is not useful anymore—because it fails to distinguish conscious objects from nonconscious things. There is nothing left to be classified as “nonconscious,” and this starkly contrasts

with the usefulness of the distinction, not only in daily life but also in clinical practice, psychology, and experimental neuroscience.

Our daily use of the term “consciousness” sometimes deviates from the sense in which it was used above, denoting phenomenal experience. If a police officer stops me at a street crossing, asking, “Did you consciously cross at that red light?” he could mean to inquire whether I was conscious of the red light when I crossed the street (phenomenal experience), or whether I made a willful decision to cross the street, knowing that the light was red. Here, consciousness is conjoined with action, and this use reflects a link with volition, planning, and decision making. This use is connoted with a distinction that has recently gained ground in the field: the dichotomy between “access consciousness” and “phenomenal consciousness.” These concepts have been introduced to distinguish processes related to generating the phenomenal content of our experiences versus processes acting upon this content and requiring access to it, such as for guidance of attention, working memory, emotional evaluation, or motor decisions and verbal expression (Block, 1990, 2005; Dehaene et al., 1998; Baars, 2002). We will return to this distinction later and examine whether it is a necessary one to make, asking whether consciousness should be divided into two kinds representing different processes or cognitive operations or whether there is only one basic process at stake (phenomenal consciousness) to which other (nonconscious) cognitive faculties need to have access.

All in all, walking through the zoo of our common vocabulary on mind and consciousness is not without peril. On the one hand, we risk making rigorous scientific analysis subordinate to folk psychology, as Paul and Patricia Churchland have extensively argued (P. S. Churchland, 1986; P. M. Churchland, 1995). By “folk psychology” we mean the set of assumptions and conceptual framework we commonly use in daily life to describe the behavior and experiences of ourselves and other people we interact with. For instance, our daily conversations are impregnated with terms like “soul” and “mind” as separable from our body, leaning on a long tradition of religiously inspired beliefs on mind and body as distinct entities that can exist independently of one another. This occurs despite the fact that most people nowadays are aware to some extent of brain research arguing how deeply intertwined and connected “mind” and “body” are. Facing the choice of engaging in a love affair with potentially adverse social consequences, a man might utter, “My mind wants something different than what my body wants.” He may well say this while conscious of the fact that “what his body wants” is as much engraved in his brain as what his mind wants, and yet continue to use this dualistic expression.

On the other hand, being suspicious of folk psychology does not mean we can dismiss the premises of our daily-life vocabulary completely. Even the most rigorous scientific analysis makes use of terms that we use in daily life. The anecdotal nature of folk psychology has sometimes been used to point out how unreliable and misguided it can be, such as when in medieval times, before the Copernican revolution, people believed the sun revolved around the earth. This unreliability should be acknowledged but does not

preempt the very concepts of “sun” and “earth” as scientifically valid constructs. Copernicus revolutionized our description of their relationship but did not abolish *these* concepts. Similarly, when we scrutinize experiences of our own or those reported by others, we will ask whether one should hold on to both “consciousness” and “brain” as concepts we regard as existing and real, and how to describe their particular relationship—using empirical evidence, like Copernicus did for the earth and sun.

1.3 Detection, Perception, and Meaning

When we try to describe a being’s interaction with its environment, the natural urge bubbles up to label its behavior as voluntary and purposeful, or reflexive and automatic. A similar question is how other beings sense their environment: should they be thought of as merely “detecting” stimuli, that is in a completely nonconscious state, or should their sensing be likened to our own, conscious perception? Recalling Aristotle, Descartes, Spinoza, Locke, and Leibniz, this question has been posed already many times throughout history and was also central to Thomas Nagel’s famous 1974 essay “What Is It Like to Be a Bat?” Imagining what it is like to be another creature can be a hilarious exercise, as illustrated in David Lodge’s (2001) novel *Thinks....* He addresses Nagel’s question by parodying how famous English writers would go about this exercise. Thinking like a bat from Salman Rushdie’s viewpoint turns out as a double-up in empathy (feeling like Salman Rushdie feeling like a bat), and the whole effort ends up “hopelessly anthropomorphic.”

“What is it like to be a thermostat?” Lodge’s lead character continues. Indeed, somewhere down the line the imagination stops: who would think of a seismograph as consciously sensing the earth’s tremor as an earthquake unfolds? Is a honeybee conscious? If we consider a motion detector in your garden that switches on a surveillance light when a cat passes in front of it, would you feel it makes sense to call this detector “conscious” of the cat’s movement? How do you come to acquire this feeling—by emotional sentiment or reasoning? You may well have trouble regarding such a simple photodetector system as conscious, as much as I do, but why is this so? Intuitively, we have trouble imagining what it would be like to be a motion detector, but a more rigorous approach will be to make a list of requirements we can reasonably associate with conscious perception and to describe which capacities a given system should have to fulfill these.

One purpose of revisiting these examples is to delineate the distinction between *detection* and *perception*. I will use the terms “detection,” “registration,” or “monitoring” whenever living beings or objects react to events transmitted to them from their environment, with the additional option that detected events may also be stored, such as by a tape recorder. This reaction need not be visible to an observer, but some evidence for a detection process is required. A webcam attached to a computer can be said to have been “detecting” when we can find the movie files it captured on the computer’s hard disk. This is a broad definition that includes cases of detection that we consider as occurring consciously, or in other

words, cases of perception. Perception is a term I will use as implying consciousness and will be interchangeable here with “sensation” or (phenomenal) experience.² The concept of “sensing” or “sensor” is often used ambiguously, sometimes referring to purely electro-mechanical detection, sometimes linked to conscious processing. I will be using this term sparingly and in a broad sense, as not implying consciousness per se but in line with the concept of “detection.” Human perception in its most exuberant form is typically cast in the visual domain. Our vision is framed in space and time, forming a whole that is presented to us in one perspective, is structured as a scene comprising distinct objects against a background, with attributes that are seamlessly integrated into each object and are highly varied and distinct, such as the color, transparency, texture, overall shape, and motion of a balloon flying in the air. However, perception may also be “simpler,” or more accurately, set in fewer sensory dimensions, such as when we smell a freshly baked apple pie with our eyes closed. Consciousness makes the essential difference between detection and perception, and one of the main hurdles in front of us is to explain what it is about the brain that makes up this difference.

No less important is the distinction between consciousness and self-consciousness. Space is lacking to even summarize the many volumes written on this topic, so I will simply outline the position taken here, which is to consider self-consciousness as a specific kind of consciousness. In chapter 8 we will look into the notion of “self” as a complex concept that is derived from the way different parts of our bodies are represented and integrated in the brain, but also from our long-term record of personal memory in association with language. This notion stands in contrast to “consciousness” as the more general case of phenomenal experience, implying that we or other beings can perceive external objects without necessarily being aware of ourselves—that is, of our own bodies or the subjective “I” we associate with our experienced thoughts, imagination, beliefs, and percepts. How brain activity relates to consciousness will thus be the basic problem in this book, whereas self-consciousness will refer to the particular case of conscious experience and representation of the self, including one’s own actions, feelings, thoughts, and percepts.

Before we turn to the actual aims of this book, I would like to make one further distinction. This concerns a widespread, dual use of the concept of “meaning.” If we are recording electrical activity from the brain and see how specifically spikes (electrical impulses) of neurons respond to external stimuli such as a sip of orange juice, the dazzling question comes up of how these neurons could ever represent something having meaning. If you recognize an image of the Eiffel Tower and attribute meaning to this sensory input, how do the neurons in your brain produce any meaningful knowledge about the image? The problem here is to find out how neuronal activity—with no obvious “meaning” by itself, and manifested by way of trains of spikes that strongly resemble a long string of zeros and ones across time—can be *about* something, can refer to the tangible and meaningful objects captured in our perception. The problem of meaning and “aboutness” has been mainly analyzed in the philosophy of language and semantics, which has coined the related

term “intentionality,” to which we will return later (Searle, 1983, 2004; see also Fodor, 1988; Putnam, 1975a,b; Burge, 1979).

Here I will argue to distinguish two different notions of meaning. Suppose we were at a gemstone exhibition, admiring a beautiful lapis lazuli, and you were to ask, “What does the color of this stone mean to you?” then I could start recalling all events, objects, attributes, and emotions associated with this dark-blue color. I would be happy reminiscing about the purchase of a ring with a lapis lazuli 21 years ago for my wife but might also think how the stone was brought from central Asian quarries to Europe, enabling Michelangelo to use dark blue colors in the Sistine Chapel’s frescoes. This sense of “meaning” is derived from all learned associations we have formed throughout our life between that color and contexts in which we perceived it. I will refer to it as the concept of “associative meaning.” The other sense of meaning refers to our phenomenal, qualitative experience of the color itself. We can logically dissociate our perception of the color from our previous daily-life experiences and learned associations. When we view the stone, we do not *have* to think of the past in order to see its color. Chapter 3 will present evidence from densely amnesic patients to examine whether perception and autobiographical memories can also be neurologically dissociated. I will refer to this second notion as “phenomenal meaning.”

Especially this second notion of meaning—the sheer qualitative nature of perceptual experience—forces us to face the problem of how our brains pull off this trick, and why. Electrophysiological recordings show that neurons in certain brain areas generate electrical activity that is specifically associated with colors. Their collective spike patterns code, or represent, particular colors—and are thus somehow *about* these colors—and a major question is how this coding arises. This book will discuss various approaches to this perplexing problem, pointing to the importance of a central coordination and multimodal integration of sensory activity in the brain, as well as the interaction between sensory and different kinds of memory systems, going beyond purely autobiographical memory.

1.4 Scope of This Book: The Brain's Representational Problem

The sheer wealth of current research in psychology and neuroscience makes it impossible to attempt a complete discussion of the topics we have already touched upon. For individual topics, an extensive literature is available in philosophy, psychology, neurology, or experimental neuroscience. A much smaller, but growing, body of work has been written on how to *connect* these two fields and, in particular, how to identify principles which can be used to bridge the gap between the domain of neural processes and the “higher realm” of conscious processing and meaningful percepts. Fueling the drive to bridge this gap (Levine, 1983), the first main aim of this book will be to compare existing theoretical approaches to neurological findings on consciousness and its relation to other cognitive functions. We will examine how different aspects of our cognitive functioning can be dissected into components considered either essential or peripheral to the problem of consciousness.

Before we embark on this in chapter 3, chapter 2 will briefly review the anatomy and functioning of the building blocks of the brain—neurons, synapses, and larger, macroscopic structures. Part of this discourse will be to dissect the central concept of “representation,” which can have very different meanings in everyday life. Photographs or paintings are regarded as representations, but how are such depictions distinguished from the information packages in our brain which we experience as rich and meaningful? Or is there no fundamental distinction between the two? David Marr (1982) defined representations as formal systems “for making explicit certain entities or types of information, together with a specification of how the system does this” (p. 20). On the one hand, this definition will be useful for isolating meaningful representations generated by brain processes from passive ones such as photographs. On the other hand, it is far from clear whether and how the brain applies formal rules in establishing representations, how entities become explicit, and why these operations would result in anything like consciousness.

In chapter 4 we will ask to what extent one of the main paradigms that has been linking neural function to cognition—computational neuroscience and, specifically, neural network modeling—has provided insights into how conscious representations may be generated by the brain. Chapter 5 will attempt to identify what may be lacking in these and other approaches, ask whether it is reasonable to consider a wealth of inanimate systems throughout nature as conscious (panpsychism), and scrutinize how we can further constrain models of conscious representation. In search of biological constraints, chapters 6 and 7 will dig into the anatomy and physiology of neural systems. Why do some of these systems engage in conscious processing, whereas others do not? What kind of physiological states of the brain are associated with consciousness—and why? Chapter 8 will integrate the various arguments and leads in the search for a representational theory on the neural basis of consciousness, outlining the requirements complex structures need to fulfill in order to give rise to conscious processing. This “wish list” will then culminate in a set of hypotheses concerning how the requirements may be implemented by neural machinery (chapter 9) and how we may conceptually bridge the gap between single-neuron processing and phenomenal experience (chapter 10). Chapter 11 will place the body of theory developed up to this point in the wider context of mind–brain philosophy and sketch some of the implications for our thinking about animal and robot consciousness and treatment of brain disorders.

A common thread throughout the book concerns the actual function of consciousness. In a nutshell, the perspective I will defend is that of an organism roaming about in a changing environment, rich in detail and perceptual complexity and often necessitating quick decisions about actions to undertake. The organism’s sensory organs continuously provide a stream of information to the brain, which then faces the task of integrating all incoming pieces into a coherent whole that can be immediately recognized, rapidly understood, and acted upon. This task is what I will call the brain’s representational problem, and consciousness is proposed as the result of the brain’s solution to this problem.

2

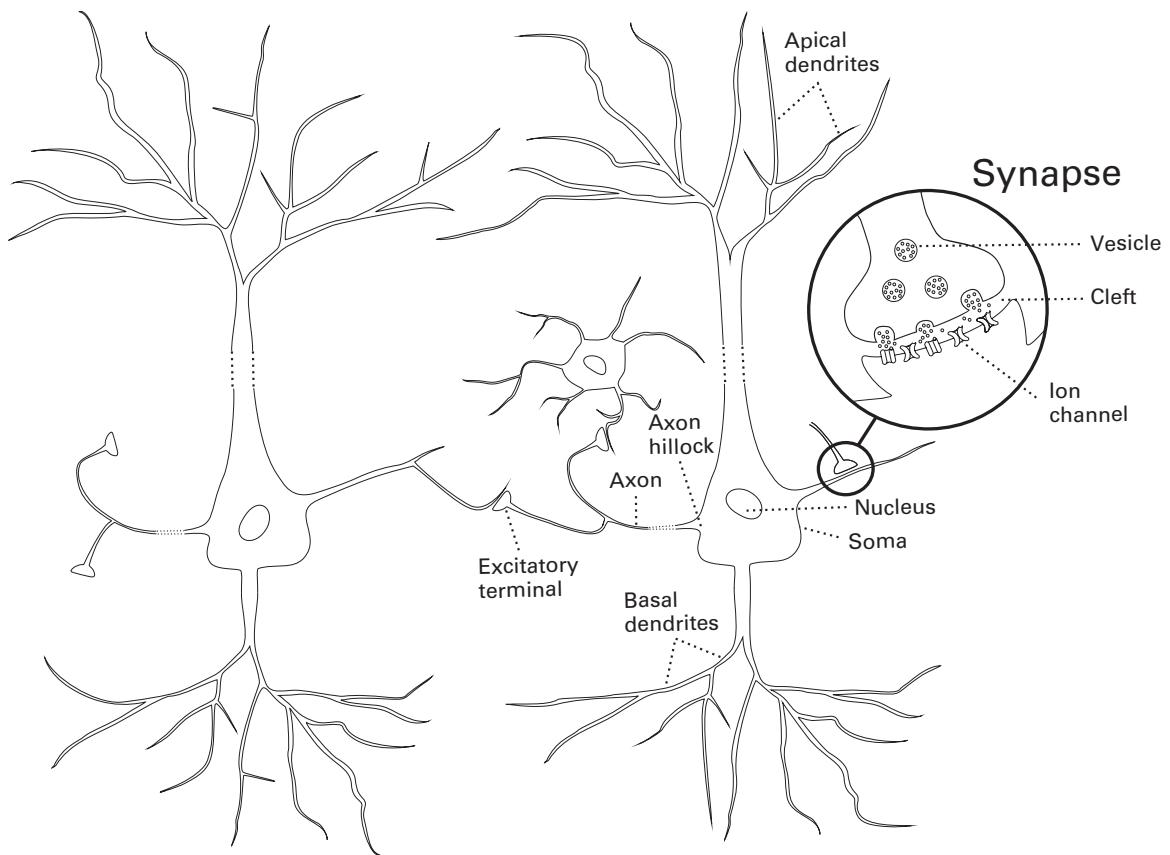
Building Representations with Neurons and Spikes

All of which instances sufficiently show that everyone judges of things according to the state of his brain, or rather mistakes for things the forms of his imagination. We need no longer wonder that there have arisen all the controversies we have witnessed, and finally skepticism: for, although human bodies in many respects agree, yet in very many others they differ [...].
—Benedict de Spinoza (1677)

2.1 Bioelectricity: The Quick and the Living

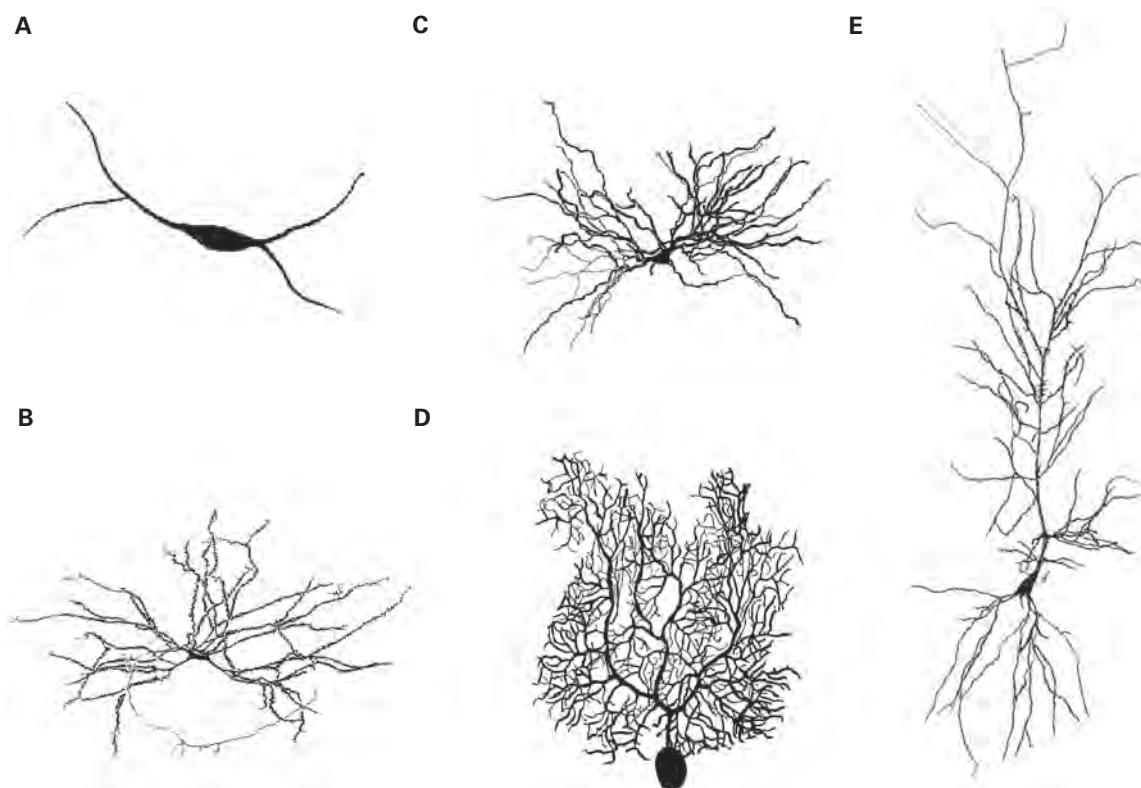
As an undergraduate student taking an introductory class in neurobiology, I was easily baffled by the sheer number of neurons in the human brain—an estimated 150 billion. By current standards, this number actually does not seem that high, for instance when compared to the U.S. national budget debt ceiling (currently about \$17,212 billion in U.S. dollars; this amounts to \$115 per neuron). Back in 1982, I was also impressed by the lightning speed at which neurons work—their spikes (also called action potentials or electrical impulses) are about 1 ms in duration, and they travel across a nerve from spinal cord to muscle at a speed up to 120 m/s. Speaking now, in the age of smartphones operating at 1.5 GHz, this neural processing speed seems far less impressive. However, taking into account that current supercomputers, despite outperforming humans at expert tasks such as chess and knowledge games such as Jeopardy, cannot integrally perform a set of daily-life tasks such as recognizing faces, loading a dishwasher, or catching a ball, the capacities of the fluffy, greasy substance inside our skulls turn out to be all the more impressive—it is fast, versatile, intelligent, flexible, resistant to damage, and capable of cognitive tasks no artificial machine has yet been shown to perform.

It is no less than remarkable that neurons, as building blocks underlying brain function, are relatively uniform in form and function. In their prototypical form they operate as a device converting a set of inputs into an output (see figure 2.1). Classically, their “receiving” or input-processing parts are the dendrites, which originate from a cell body or soma and ramify into finer and finer branches, forming tree-like patterns characteristic of cell types and specific brain structures (see figure 2.2). On average, a neuron receives about

**Figure 2.1**

Scheme of two connected neurons, modeled after pyramidal cells in the cortex. Inset shows the core elements of a synapse: presynaptic vesicles filled with transmitter molecules, synaptic cleft and ion channels in the postsynaptic membrane.

2,000 inputs from other cells, and the vast majority of these inputs arrive at contact points called synapses. Synapses are the sites where much of the cell's action occurs: where most drugs find their molecular targets and where much of a neuron's flexibility or malleability (plasticity) is located. Consisting of a presynaptic and postsynaptic part, separated by a narrow cleft, the critical event at a conventional synapse is the release of a chemical compound, a neurotransmitter, from the presynaptic part into the cleft. Being only about 30 nm wide, the cleft forms a narrow space that transmitter molecules diffuse across to bind to receptor proteins in the postsynaptic membrane within a millisecond. In most synapses this binding causes an electric current to arise across the local postsynaptic membrane, and the accompanying change in voltage across the membrane is then propagated along the

**Figure 2.2**

Examples of different neuron types found throughout the central nervous system. (A) Neuron in the basal forebrain, sensitive to blood pressure. (B) Projection neuron of the striatum, characterized by spines on dendrites. (C) Relay neuron in the posterior thalamus. (D) Cerebellar Purkinje cell, characterized by its large dendritic tree laid out in a two-dimensional plane. (E) Pyramidal cell in the hippocampus, with its apical dendrites pointing upward and basal dendrites downward. Example neurons are of different sizes and are not drawn on the same scale. Adapted from Kirouac and Pittman (1999; A), Kita et al. (1984; B), Li et al. (2003; C), Roth and Häusser (2001; D), and Losavio et al. (2008; E).

dendrite and toward the soma. The soma is the site where all inputs received on the dendrites are integrated, but it also harbors inside the nucleus, most of the cell's protein-synthesizing machinery, and an elaborate trafficking system for transporting proteins and messenger ribonucleic acid to remote locations in axons or dendrites. A specialized zone of the somatic membrane—the *axon hillock*—is where the axon originates, a very thin fiber, frequently branching into daughter axons (collaterals) and rapidly propagating the cell's electrical impulses to target cells. Once the impulse arrives at the terminal of an axon, which forms the presynaptic part of a synapse onto its target neuron, it unleashes a cascade of electrochemical events resulting in the release of neurotransmitter molecules into the synaptic cleft contacting the next neuron in a chain of connections (see figure 2.1).

This is the classic picture of a neuron. Taking neurons from different brain structures altogether and enlarging them to human proportions, more than a big zoo could be filled with all their varieties. In the giraffe-like compounds we would find the pyramidal cells of the hippocampus with their large apical dendrites (see figure 2.2), whereas a bipolar cell from the retina would be the size of a cockroach. This heterogeneity is mirrored by a large variety of neurotransmitters, of which more than a hundred have been identified, leaving hormones and other neurochemicals diffusing throughout the wider space surrounding neurons aside. Two major types of neurotransmitter can be distinguished: ionotropic and metabotropic. The ionotropic transmitters, including glutamate, gamma-aminobutyric acid (GABA), and acetylcholine, bind to a receptor protein located in the postsynaptic membrane, and this receptor forms a molecular pore or channel that can be in a closed or open state, so that ions of a particular kind (e.g., Na^+ or K^+) in the intra- or extracellular fluid can either pass through or not. Ion channels typically open up when having bound a transmitter molecule, but for some types the opening state is subject to additional conditions—for instance, the voltage difference that exists across the membrane. Ionotropic transmitters and their receptor channels act as the fast workhorses of the nervous system: their ionic currents and associated voltage changes last in the order of a few up to tens of milliseconds. Metabotropic transmitters bind to receptors that trigger a much slower cascade of biochemical (“metabolic”) reactions in the postsynaptic neuron. As a consequence, a wide range of cellular functions can be affected via the synthesis of intracellular messenger molecules and subsequent biochemical steps. The opening and closing properties of ion channels can be altered, protein functions can be modified, and even gene transcription can be regulated, causing long-lasting changes in protein metabolism. If ionotropic transmitters and their receptor channels are epitomized as fast workhorses, or *drivers* of information transmission, then metabotropic transmitters act as slower *modulators* or “state setters.” For instance, they tweak the number of excitatory inputs that are required to make the postsynaptic neuron fire. They, too, come in various shapes and sizes, ranging from monoamine transmitters such as dopamine and serotonin to the less conspicuous neuropeptides, which are short, snake-like chains of amino acids including the ill-famed opioid peptides endogenous to our brain.

If an ion channel opens up, ions of a certain type (e.g., Na^+) will flow through it. This flow is possible, first of all, because the concentrations of Na^+ on the inside and outside of the neuron are different; there is a chemical gradient resulting in diffusion. Secondly, the electric voltage is also different between the interior and exterior surface of the neuron. An electric gradient exists across the membrane, with the inside of the cell having a negative potential relative to the outside, causing Na^+ ions to rush inward. Both gradients can be at play at the same time, giving rise to the electrodiffusion theory developed by Nernst in the late nineteenth century. The direction and extent to which ions of a given species flow through a membrane channel are determined by the net balance of the chemical gradient of that species and the voltage across the membrane. A channel is considered selective if

only one ionic species passes through it, but throughout the brain we find many examples of nonselective channels.

Under resting conditions, the voltage of the cell's interior is negative relative to the extracellular space and lies around -65 mV . The cause of this resting-state negativity lies primarily in the higher concentration of K^+ ions inside the cell than on the outside, and during rest, specific types of K^+ -selective channels remain in a relatively "open" state (as compared to other channel types, selective for Na^+ and Cl^-): a modest flux of K^+ ions will continuously leak to the cell's exterior. This continuous lack of net positive charge corresponds to the negative voltage difference across the membrane. The voltage gradient is referred to as a state of *polarization*, which in practice means a *negative* polarization of the cell's interior relative to the exterior. A depolarization denotes a decrement in polarization (when membrane voltage moves from a strongly negative potential to a less negative range), whereas a hyperpolarization occurs when the polarization increases.

What happens if, in the end, so many K^+ ions have left the cell that the gradient is extinguished? This would result in a loss of polarization were it not for the ingenuity of ionic pumps in the membrane that are quietly working in the background. These energy-consuming proteins work as vehicles that shuttle multiple types of ions to the cell's interior or exterior. After a barrage of intensive neural activity, pumps will work hard to transport K^+ ions back into the cell, in exchange for Na^+ ions that are expelled to the exterior. Much of the brain's energy—adding up to 25% of the body's total glucose-fuel demands—is spent on this slow, ongoing maintenance of electrochemical gradients.

Unlike the continuous stream of K^+ ions seeping out of the neuron at rest, a spike is made up of an ultrafast opening of Na^+ channels. This happens when the neuron's membrane is sufficiently depolarized, usually because the sum of its synaptic inputs has become large enough. A neuron generating a spike is also said to "fire" (indeed an audio signal reflecting the membrane voltage of a repetitively firing neuron sounds much like a machine gun). The value of membrane voltage where these Na^+ channels open up—first in small numbers but then rapidly and massively—is called the spike threshold. As Na^+ ions pour into the cell, their effect on the membrane voltage is to depolarize it even further because the influx of positive charge diminishes the excess negative charge inside the cell dominating the resting voltage. In less than a millisecond, a positive feedback loop arises: the greater the depolarization, the more Na^+ channels open and the stronger the influx of Na^+ ions. Once the loop is ignited, there is no way for the cell to "undo" the cascade of events: the spike is an all-or-none signal. The membrane potential briefly shoots up to a level of reversed polarity, with the cell's interior becoming positive.

Next, the cell's molecular machinery works to bring the neuron back to rest. The first step in this restoration of order is that the high level of depolarization causes the Na^+ pore to close. Meanwhile, the strong depolarization also causes the opening of a specialized type of K^+ channel. Once this current is unleashed, K^+ ions flow from the cell's inside to the outside, restoring the cell's excess of negative charge and bringing back the membrane

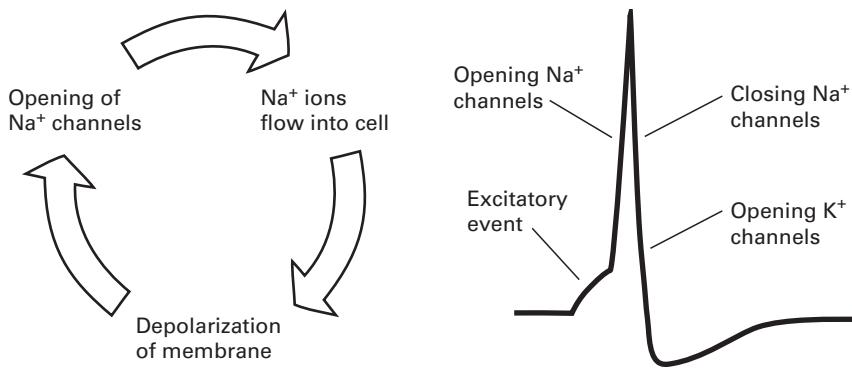


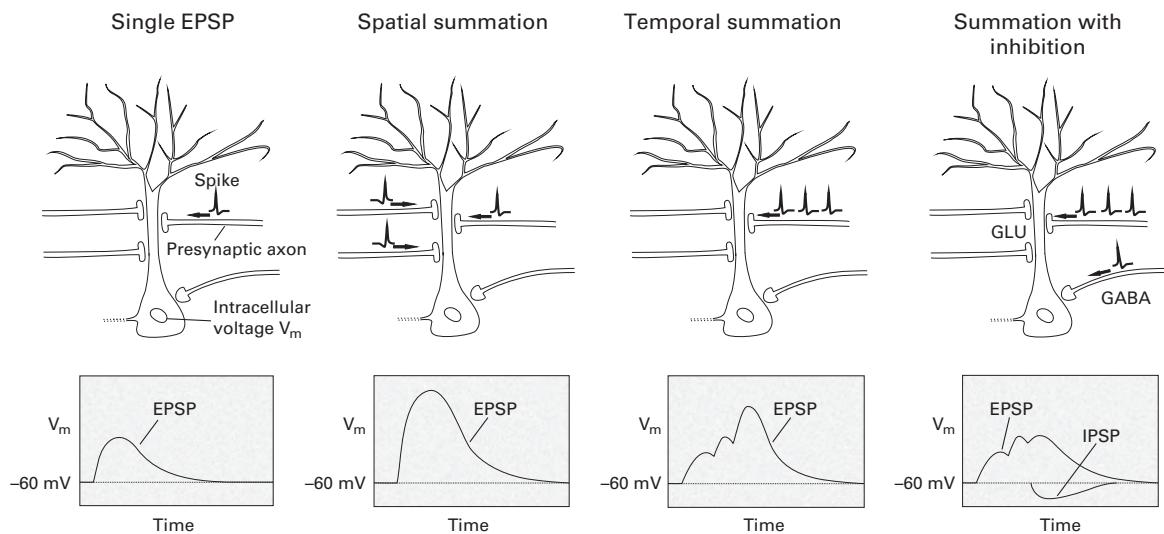
Figure 2.3

Cycle of electrophysiological events leading to the firing of an action potential. Left: An initial depolarization of the membrane results in opening of Na^+ channels and influx of ions into the cell, which will further enhance the depolarization (positive feedback loop). Right: Time course of the action potential. As a result of an excitatory input or other event, Na^+ channels open massively during the rising phase of the action potential. During the falling (repolarizing) phase, these channels close whereas K^+ channels open. The dip below the horizontal line (baseline or resting potential) is called the spike afterhyperpolarization.

potential to its resting range. More events than this interplay of Na^+ and K^+ currents take place during spike generation, but this captures the basic biophysics of how spikes—as unitary pieces of information—are generated (see figure 2.3).

Ion channels are exquisitely sensitive to chemical substances such as neurotoxins. *Fugu*, or puffer fish, is served in Japanese restaurants as a delicacy and causes a few victims each year, mainly in the domestic cuisine. It contains symbiotic bacteria that produce the extremely powerful Na^+ channel blocker tetrodotoxin. Also the local anesthetics lidocaine and novocaine belong to the family of drugs affecting spike-mediating Na^+ currents. Less deadly than tetrodotoxin, their use in dental practice aptly illustrates how vital Na^+ channels, whether in your brain or in peripheral nerve branches, are for conscious sensations to arise.

Returning to the brain's fast workhorses, how do ionotropic transmitters induce changes in the cell's polarization state, and how do they influence whether a cell will fire or not? We can distinguish excitatory (spike-promoting) transmitters, such as glutamate, versus inhibitory (spike-suppressing) transmitters, such as GABA. Although the amino acid glutamate is abundant in both our brains and food, it escaped detection as the brain's main excitatory chemical until the late 1970s and early 1980s, when Watkins, Evans, Collingridge, Cotman, Lynch, and colleagues used pharmacological receptor blockers to elucidate the role of glutamate-binding receptors in the brain (Watkins & Evans, 1981). When glutamate molecules are released into the synaptic cleft, they bind to a variety of proteins in the post-synaptic membrane, of which the AMPA receptors (named after an artificially synthesized drug) stand out as the most important ones. Once the AMPA receptor binds glutamate, its

**Figure 2.4**

Processing of single or multiple synaptic inputs by neurons. A single presynaptic spike (leftmost panel) elicits a small excitatory postsynaptic potential (EPSP) in the postsynaptic cell. In the case of spatial summation, multiple presynaptic fibers terminating on the neuron of interest are activated. When spikes arrive simultaneously, their individual EPSPs add up to a larger potential than that resulting from a single-input EPSP. In the case of temporal summation, spikes arriving within milliseconds after one another along the same fiber elicit EPSPs that will overlap in time, resulting in a gradually rising, compound EPSP. As illustrated in the rightmost panel, fast excitatory inputs in the brain are often mediated by glutamatergic (GLU) synapses. When the third excitatory input, temporally summing with the previous two, coincides in time with a GABAergic input, the peak of the compound EPSP is reduced due to the inhibitory effect of GABA on the postsynaptic cell. In this scheme, EPSPs are small in amplitude and do not give rise to spikes.

ionic pore opens, mainly letting through Na^+ ions from the outside into the cell. If the cell receives this input under resting conditions, the net result is a fast depolarization.

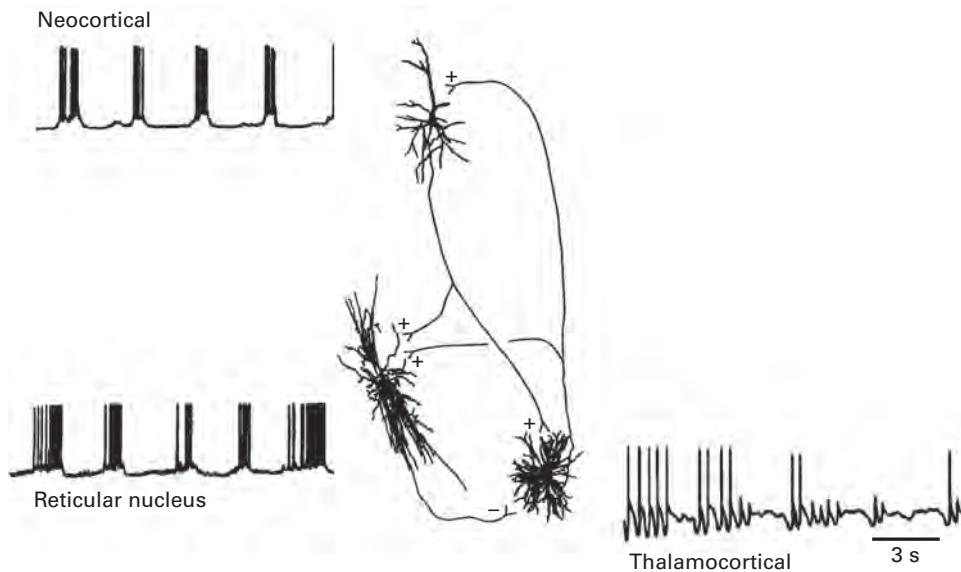
A single synaptic input will not be strong enough to drive the cell's membrane potential up to spike threshold, but if a sufficient number of fellow synapses are active around the same time, then a *summation* of inputs will take place because they overlap in time (see figure 2.4). Now, the membrane potential may reach the threshold so that a spike event is ignited. The kind of potential change caused by a glutamatergic input is called an *excitatory postsynaptic potential* (EPSP). A pyramidal cell needs in the order of tens of excitatory inputs to reach threshold (Brown et al., 1979). How could it ever fire when, seemingly by chance, so many inputs need to coincide in a time window only about 25 ms wide? Part of the explanation lies in the sheer number (~2,000) of synaptic contacts a neuron receives. Another significant fact is that many neurons in the awake or sleeping brain have a background activity—they fire spikes usually at a low rate. In many brain regions, this background spiking is ascribed to excitatory synaptic inputs a neuron receives from other active neurons, and it contributes itself to background activity of other neurons. Finally,

summation of temporally overlapping inputs can turn out stronger than expected by chance because the activity of presynaptic neurons may be coordinated in time. Such coordination may easily arise when these presynaptic neurons are themselves interconnected.

The spike output of a neuron is sculpted in part by the coincident pattern of excitatory inputs, but another determinant is its inhibitory input, primarily mediated by GABA. GABA commonly binds to postsynaptic receptor channels that selectively pass Cl^- ions. The effect of the resulting Cl^- flux is not so much to cause an unconditional hyperpolarization of the postsynaptic neuron but to *shunt* the membrane, lowering its resistance. By way of Ohm's law, a lowered membrane resistance will cause another, excitatory input to have less impact on the membrane potential. Hence, this chloride-mediated potential is dubbed an inhibitory postsynaptic potential (IPSP), which is usually justified. In most types of neuron, the shunting effect of GABA will be accompanied by membrane hyperpolarization, but in exceptional circumstances (e.g., in the newborn brain) GABA induces a depolarization which may even promote spikes (Cherubini et al., 1991; De Jeu & Pennartz, 2002).

Ion channels and transmitters thus make up the basic repertoire that our neuronal orchestras use to play their peculiar kind of music. But as in real music, one's attention is especially captured by the variations on a theme—by the surprises and exceptions. The last 3 decades have elucidated many exceptions to the standard model of the neuron—and these shed light on the functional versatility of neuronal networks. For instance, several brain nuclei contain cells that generate spikes even in the absence of excitatory inputs. Their basal state is sufficiently depolarized to allow the spontaneous generation of repetitive spike patterns—where “spontaneous” means that spikes arise even in the absence of synaptic inputs. This occurs, for example, in the suprachiasmatic nucleus—a tiny hypothalamic nucleus harboring the brain's biological day–night clock—and in the dopaminergic cells of the mesencephalon, which are dysfunctional in Parkinson's disease (Inouye & Kawamura, 1979; Grace & Bunney, 1984; Pennartz, de Jeu et al., 2002). A second exceptional type of cell is the thalamic relay neuron, which processes sensory information originating in the peripheral senses and transmits its output to the neocortex (see figure 2.5). In the visual system, relay cells in the lateral geniculate nucleus (LGN) of the thalamus receive information via the optic nerve from the retina and send their outputs to the primary visual cortex (area V1).

If a thalamic relay cell hovers in a depolarized state but also receives GABAergic input from a neighboring thalamic region (the reticular nucleus), this input will briefly shunt and hyperpolarize the relay cell's membrane. At first, this will prevent the relay cell from firing, but now the GABAergic inhibition comes with a vengeance. Because of the hyperpolarization, a specific type of Ca^{2+} channel in the same membrane assumes a state in which its opening can be easily triggered once the inhibition fades out. Now, the membrane voltage climbs back to resting level and Ca^{2+} channels open, providing a rapid flux of positive charge into the cell, leading up to a low-threshold Ca^{2+} spike. This slow event triggers a series of fast Na^+ spikes on top of it. As an end result, the circuit composed of

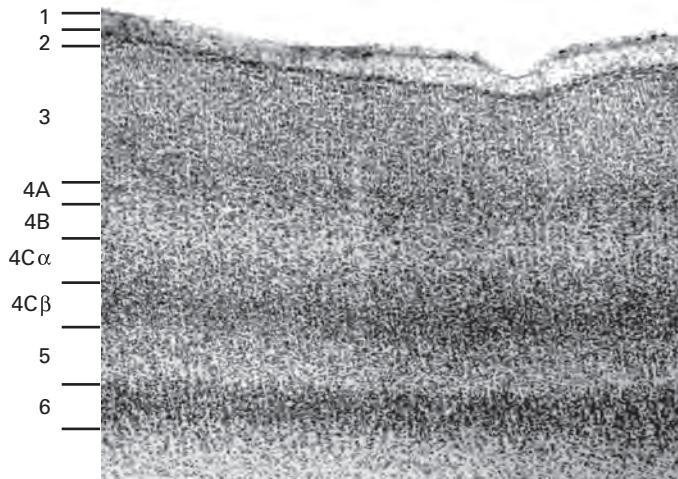
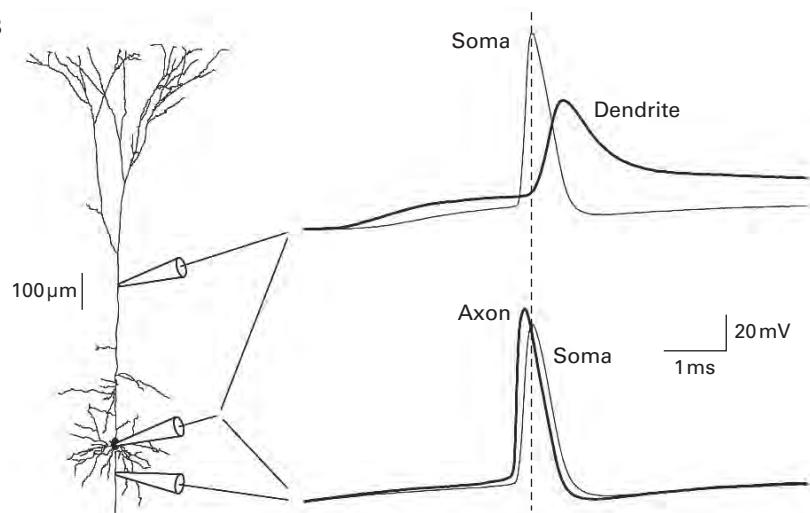
**Figure 2.5**

Circuit diagram with a thalamocortical cell (lower right: relay neuron), neocortical cell (top), and cell in the reticular nucleus of the thalamus (left). Accompanying traces show transmembrane voltage patterns across time. The excitatory and inhibitory nature of synaptic connections is indicated by "+" and "-" signs, respectively. The membrane potential patterns shown here are characteristic of a brain state of deep sleep and can also be found under anesthesia. Note that the spikes of the thalamocortical cell are preceded in time by transient hyperpolarizations, likely corresponding to GABAergic inputs. Adapted from Steriade, McCormick, & Sejnowski (1993).

thalamic relay cells, neocortex, and reticular nucleus generates rhythmic activity made up of a silent, hyperpolarized phase of an oscillatory cycle and an active, depolarized phase. Again, GABA is not merely acting as an inhibitory transmitter.

2.2 From Single Cells to Circuits: The Neocortex as an Example

The neocortex has been introduced as forming a single element connected with thalamic relay nuclei, whereas it is composed of multiple layers, each with specific inputs and outputs and characteristic cell types. In primates, the neocortical sheet is folded in *gyri* (hills, protrusions) alternating with *sulci* (valleys). In many “lower” vertebrates such as mice and rats this folding is absent. Physiologists and anatomists have classically distinguished six layers, but in many cases sublayers have been outlined, and in particular cortical regions some layers are only poorly visible or even absent (see figure 2.6; an example is the anterior cingulate cortex, where layer IV, abbreviated here as L4, is lacking). Starting just beneath the skull, L1 is largely devoid of cell bodies and is filled with a dense meshwork of axons. Many of these fibers originate from the superficial (or upper) layers 2 and 3, which contain mostly small cell bodies with modest dendritic trees.

A**B****Figure 2.6**

(A) Section of monkey primary visual cortex (V1), histologically stained to visualize the layers (laminae) of the neocortex (numbers indicated on left-hand side). Layer 1 (L1) marks the pial surface of the cortex. Note how L4 in primates is subdivided into several sublayers. Adapted from Saul et al. (2005). (B) L5 harbors many pyramidal cells with apical dendrites extending a long way toward the pial surface. If action potentials generated by the cell are recorded with three fine glass pipettes, it can be observed that they first arise in the axon hillock and axon, a fraction of a millisecond later in the soma, and then propagate into the apical dendrites (right-hand panel). From Stuart et al. (1997), with permission from Elsevier.

Deeper down lies L4, again densely populated with neurons, which are the recipients of the output from thalamic relay cells. In primary sensory areas of the neocortex, the main recipient is the spiny stellate cell, named after its star-like dendritic branching pattern and tiny, thorn-like protrusions on the dendrites (spines), where much of the cell's synaptic contacts are made. The deep layers 5 and 6 harbor the largest neurons of the neocortex: cone- or pyramid-shaped cell bodies with their apex pointing toward the skull. These pyramidal cells send up a huge dendrite, with branches ramifying into the superficial layers. Emerging from the bottom of the pyramid, the basal dendrites branch into the deep layers. L5–6 neurons are considered cellular processing elements that integrate inputs from other cortical areas and synthesize an output that is broadcast to subcortical areas, including the thalamus.

Interspersed across the six layers are a wealth of interneurons. By definition, these cells do not send their axons to remote cortical areas or subcortical structures but project locally. Cortical cells were first described in the late nineteenth and early twentieth centuries by Santiago Ramón y Cajal, Camillo Golgi, and their colleagues, and at that time silver impregnations and other coarse staining techniques were the only methods available. Initial classifications were based merely on morphology, often reminiscent of the upper bourgeoisie enjoying the *Belle Époque* that preceded the First World War: chandelier cells, double-bouquet cells, and basket cells. Now, about a century later, many more different classes and varieties of interneurons have been recognized in the neocortex alone (Ascoli et al., 2008), but a common hallmark is that they use GABA as transmitter. The classification is based partly on specific neurotransmitters and receptors that colocalize with GABA in the same cell type and partly on the connectivity between these neurons and other cell types.

A striking consequence of the spatial and connectional arrangement of neocortical neurons is that they generate electrical mass signals, which can be measured inside the extracellular space of the cortex but also on the outside, even on the human scalp. Strong electroencephalographic (EEG) signals can be measured especially when they originate from cell groups having their somata lying next to each other, their dendrites neatly aligned in the same orientation. If a group of cortical cells in this configuration receives a uniform excitatory input, the EPSP activity observable inside the cells will be paralleled by a collective loss of positive charge on their outside, observable as a negative voltage deflection on an electrode placed in extracellular space. These voltage deflections can be measured on the upper surface of the cortex. When we record EEG activity that may be assumed to originate locally in brain tissue surrounding an electrode, it is referred to as the local field potential (LFP). Since the initial discovery of periodic EEG activity in humans by Hans Berger reported in 1929, many types of brain rhythm have been described, of which the main frequency bands are alpha (8–13 Hz), beta (14–30 Hz), gamma (30–100 Hz), delta (1–4 Hz), and theta (4–8 Hz; these frequency bands can be defined differently throughout the literature, depending for instance on the species studied). The neural mechanisms generating these rhythms are not fully understood, but at least theta and gamma rhythmicity

are known to consist of cycles of (glutamatergic) excitation and (GABAergic) inhibition, mediated via synaptic loops connecting pyramidal cells and interneurons (Wang & Buzsaki, 1996; Rotstein et al., 2005; Buzsaki, 2006).

Set against the backdrop of this neocortical jungle of neurons with their extensive lianas, Greg Stuart and Bert Sakmann (1994) discovered a significant variation on the “standard” model of the neuron (see figure 2.6B). When they elicited a spike in an L5 neuron, via an electrode placed at the cell body, and measured how that spike affected the membrane voltage way up the apical dendrite, they found that this spike was actively propagated from the soma *up* the apical dendrite. This meant a violation of the classic rule stating that traffic along the dendrites should only go from the remote synapses toward the soma and not in the other direction. This oddly directed spike is called a back-propagating action potential. Once a spike arrives in superficially located branch points of a cortical cell’s dendritic arbor, it may coincide with EPSPs arising from synaptic contacts onto the same apical dendritic branches. This coincidence leads to an unusual, strong amplification of the synaptic input by way of calcium current. In this way, back-propagating dendritic spikes and calcium current provide a mechanism for associating inputs from deep and superficial layers (Larkum et al., 1999).

2.3 Synaptic Plasticity: Neuronal Records of Experience

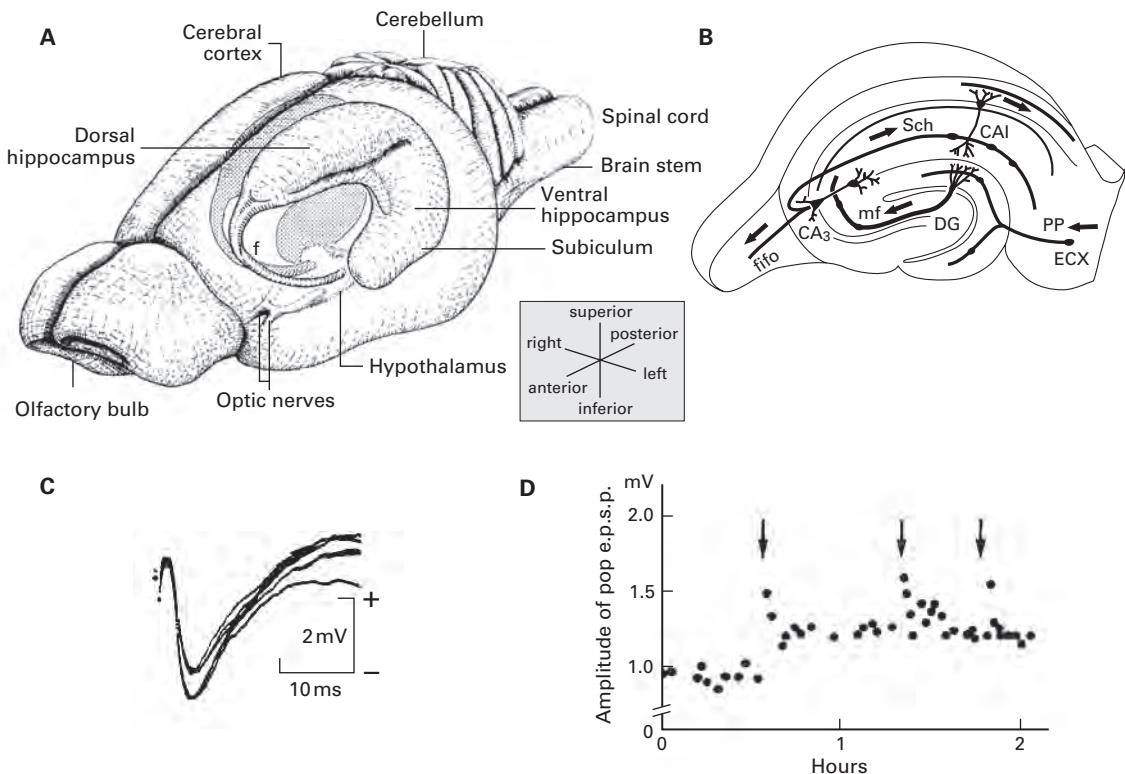
As an experience-dependent, lasting adaptation to events happening in our environment and inside our bodies, learning has been studied throughout the animal kingdom, ranging from sea snails and leeches to primates. Working in the 1970s with neural circuits that mediate reflexes in the sea hare *Aplysia californica*, Kandel and colleagues discovered that adaptations of the gill-withdrawal response to an object touching the animal’s siphon—a spout by which the animal squirts out debris and seawater—depend on changes in the strength of synaptic transmission in these circuits (Castellucci et al., 1978). A stronger reflex response was coupled to an enhancement of synaptic connections between neurons sensitive to siphon touches and neurons commanding the retraction of the snail’s gill, as a protection against potential threat. When the reflex was weakened by repeating the siphon touch without any adverse or beneficial consequences for the animal, a downgrading of the synaptic connection was observed.

By now, this adaptive ability of synapses to go up or down in strength—synaptic plasticity—has been described in numerous animal species and within many brain structures. Well before the *Aplysia* experiments on synaptic plasticity, Bliss and Lømo (1973; Bliss & Gardner-Medwin, 1973) had discovered that fast, repetitive electrical stimulation of fiber pathways in the hippocampus of the rabbit induces a rapid enhancement of synaptic strength, lasting for at least several hours (see figure 2.7). This form of synaptic plasticity—long-term potentiation (LTP)—became the most widely studied cellular model of learning and memory. LTP and its counterpart, long-term depression (LTD), have been found in a

wealth of vertebrate brain structures. This nearly ubiquitous occurrence throughout the central nervous system already hints at how *basic* synaptic plasticity actually is—it underlies most forms of nervous adaptiveness in general, not exclusively the “higher” forms of memory that we associate with recollections from our daily life. The rich, detailed memories about an eventful holiday that we can retrieve after returning home are known to be associated with brain structures in our temporal lobe (in particular the hippocampus) but are believed to represent only one, highly evolved kind of memory function that relies on synaptic plasticity.

The glutamatergic synapse presents us with a well-studied case of plasticity of fast neurotransmission. In addition to AMPA receptors, glutamate released from a presynaptic terminal will bind to another type of ionotropic channel, the NMDA receptor (named after a chemical that specifically binds to it, *N*-methyl-D-aspartate). The NMDA receptor is a masterpiece of molecular engineering: it works as a detector of *coincident* pre- and postsynaptic activity. Presynaptic activity is manifested by the release of glutamate. Glutamate binds to both AMPA and NMDA receptor channels, but only the ion channel of the AMPA receptor will unconditionally open in response to this binding. The NMDA receptor is more demanding. The opening of its channel, which is permeable to Ca^{2+} ions, also requires that the voltage across the patch of dendritic membrane harboring the receptor is sufficiently depolarized, that is, to a range of about -45 mV or even less negative. This depolarizing demand arises because, in resting conditions, a Mg^{2+} ion is situated within the channel toward its extracellular opening and blocks the passage of cations. A highly negative voltage inside the cell is sufficient to keep this Mg^{2+} ion in its den. However, upon a strong depolarization of the cell, the ion flees into extracellular space, allowing Ca^{2+} ions to enter the cell. The opening of NMDA receptor channels causes additional membrane depolarization, on top of the initial depolarization already supplied by AMPA receptors. Once the Ca^{2+} ions are inside the dendrite, they bring about a biochemical chain of events resulting in a long-lasting strengthening or weakening of the synapse. In his book *The Organization of Behavior* (1949), the great Canadian psychologist Donald Hebb theorized that associative learning is mediated by changes in synaptic strength that are determined by the activity state of both the pre- and postsynaptic element, and hence this type of plasticity is called “Hebbian.”

Several molecular mechanisms to bring about a semi-permanent change of synaptic strength have been elucidated: the number of ready-to-use AMPA receptors in the postsynaptic membrane is regulated via trafficking mechanisms, or the functioning of single AMPA receptors, already in the membrane, can be modified by adding or removing phosphate residues (Malenka & Nicoll, 1999). Recently, even more long-lasting, structural changes in synapses have received attention. The spine, as a postsynaptic part of the synapse, may undergo a widening or shrinkage of receptive surface area or even a splitting of one receptive zone into two (Luscher et al., 2000). An even more extreme, binary form of synaptic plasticity is to regulate whether or not a synapse between two neurons is established in the

**Figure 2.7**

(A) Location of the hippocampus relative to other structures in the rat brain. In the denotation of the anatomical coordinate system on the lower right, “superior” is equivalent to “dorsal” and “inferior” to ventral. f, fornix. From Amaral and Witter (1995). (B) If the hippocampus is cut into slices that lie orthogonal to the main axis running from the dorsal to ventral hippocampus, a lamellar organization is revealed. In each slice, several subregions can be distinguished: CA1, cornu ammonis field 1; CA3, cornu ammonis field 3; DG, dentate gyrus; ECx, entorhinal cortex. These regions are connected by several pathways consisting of axonal fibers: PP, perforant path; mf, mossy fibers; fifo, fimbria-fornix; Sch, Schaffer collaterals. Arrows indicate direction of spike propagation. Adapted from Andersen et al. (1969). (C) Waveforms reflecting excitatory postsynaptic potentials (EPSPs) as recorded extracellularly. These field recordings were made in the dentate gyrus, and the responses were elicited by electrically stimulating the perforant path in the anesthetized rabbit. The largest EPSPs are recorded after the perforant path was stimulated by a train of electric pulses repeated at a high rate (100 per second for 4 seconds). This so-called “tetanus” induces a long-lasting enhancement of synaptic strength (long-term potentiation). (D) Time course of the amplitude of the field EPSP (“pop e.p.s.p.”) across 2 hours; tetani are indicated by downward arrows. From Bliss and Lømo (1973). (A, C, D): reproduced with permission (copyright on A: Elsevier; copyright on C and D: Wiley).

first place. The recent technique of *in vivo* two-photon imaging has enabled researchers to track the experience-dependent genesis or elimination of synaptic contacts over prolonged periods of time (Trachtenberg et al., 2002; Holtmaat et al., 2006).

Is synaptic plasticity indeed causally important for learning and memory in vertebrates? A pharmacological blockade of NMDA receptors in the hippocampus was shown to impair induction of LTP in this structure as well as spatial learning (Morris et al., 1986). In rodents, learning about fearful stimuli is widely studied by pairing a tone with a subsequent electric shock. This basic form of learning is expressed behaviorally by the animal's beginning to "freeze" (immobilize) when hearing the tone, whereas prior to learning this freezing behavior was only displayed when receiving the shock. After a couple of tone–shock pairings, spike rates of amygdala neurons that respond to the tone, and concurrent local field responses reflecting their synaptic input, rise in parallel with behavioral learning, and this response enhancement is sensitive to NMDA receptor blockade (Rogan et al., 1997; Maren & Quirk, 2004).

The discovery of LTP and related forms of synaptic plasticity is often brought up as a success story, and with good reason. Of all cognitive processes, the neural substrates of learning and long-term memory are arguably known best. However, this success should not distract us from the fact of how basic our empirical knowledge of this neural substrate still is. Our view has been confined to studying one type of synapse within a more complex network of brain structures at a time, and often the experimental treatment has been to subject synapses of a given type to a uniform manipulation, such as massive repetitive electrical stimulation. We still know extremely little about the more varied, distributed microchanges occurring in neuronal networks operating under natural conditions, when, for instance, storing a picture of a face or a jazz melody. Our knowledge does not go much further than understanding synaptic pathways as mediating simple input–output reactions (comparable to reflexes), even if these are thought to be part of systems mediating higher cognitive functions, such as autobiographic memory. But cellular and molecular as much of this knowledge may be, it has already been tremendously instructive for the construction of computational models of these more complex forms of memory.

2.4 A Bird's Eye View of the Mammalian Brain: From Medulla to Diencephalon

We will now step from the aggregate level of synapses and neurons to the higher level of brain structures, surveying some key anatomical divisions of the central nervous system. Neuroanatomy, in my experience, is not just basic and essential but can also be fascinating and puzzling. Following the architect Louis Sullivan's *adagium* "form follows function," it provides a wealth of clues as to how sensory, motor, and cognitive functions are accomplished. If we survey the seven main divisions of the central nervous system across vertebrate species (see figure 2.8A)—the spinal cord, medulla, pons, cerebellum, mesencephalon (or midbrain), diencephalon, and telencephalon—it is striking how relatively

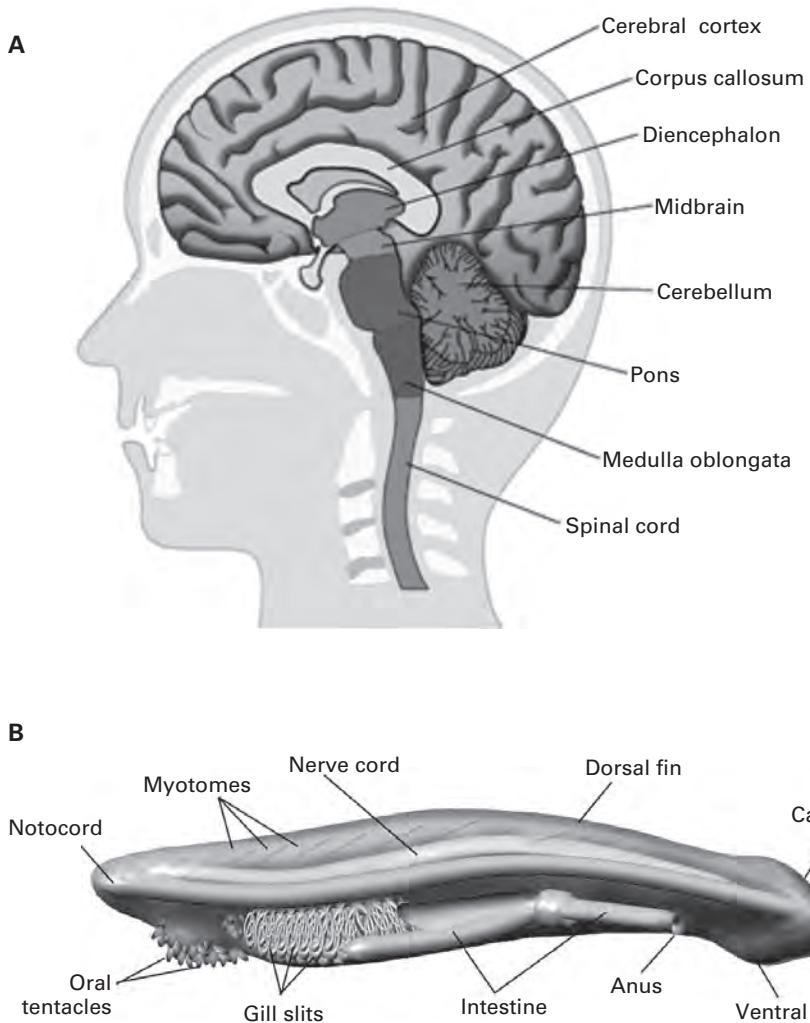


Figure 2.8

(A) Seven main divisions of the human brain. Medial view of the brain and upper part of the spinal cord. The cerebral cortex (neocortex) and corpus callosum are part of the telencephalon, which comprises more structures than are shown here. Based on Nieuwenhuys et al. (1988). (B) Overview of the internal anatomy of the lancelet, *Amphioxus*. Although spineless, the lancelet possesses a nerve cord analogous to the spinal cord in vertebrates. Rigidity is provided by the notochord, consisting of elastic tissue. The lancelet does have oral tentacles for feeding but lacks a distinct brain-like expansion of the anterior end of the nerve cord. From Benito-Gutiérrez (2006; copyright IIVSspring).

conserved the “lower” parts such as spinal cord and medulla are across species, and how varied and extravagant the evolutionary development of the telencephalon, including the cerebral cortex, has been. This evolutionary trend can be understood from the increasing processing and control demands arising when more complex vertebrates evolved from “simple” chordate species, still represented today by the lancelet, *Amphioxus* (see figure 2.8B). Despite this diversification, the basic neuroanatomic design is very similar among different mammalian species such as rats, monkeys, and humans, and findings from these species will be treated here on an equal footing.

As in *Homo sapiens*, the design of the nervous cord in spineless creatures like the lancelet can be cast as serving two basic functions: processing sensory information from the skin and muscles of the trunk and converting this into reactive motor commands. This captures the basics of a reflex arc: if your skin touches a sharp, pointed object, immediately activate those muscles that will make you retract. That these creatures are capable of directed swimming implies that only one of the two cords’ endings faces the direction of forward movement. This end is the one that is confronted first in time with the most novel, interesting information about the environment. Useful as it might be to have a pair of extra eyes and ears at your tail (if you were a lancelet), it is better to have them at a body location where you are facing the most unknown part of the outer world: the part where you are heading. This frontal end also happens to be the location where the endodermal feeding tube (or gastrointestinal tract) opens up, making it the spot where the development of jaws, teeth, and organs for smell and taste is most useful. This provides the evolutionary rationale for cephalization and the formation of brain-like frontal nervous buds, developing into full-blown heads with brains, jaws, eyes, and ears along the evolution of vertebrates.

It thus makes tremendous evolutionary sense to equip the nervous system with a brain next to a spinal cord, and equip this expanding bud with cranial nerves and visual, auditory, vestibular, and chemical sensors. Also the muscles controlling eye movement, jaws, tongue, neck, and face need to be controlled, requiring transmission of motor commands from the brain toward all of this cephalic equipment. These command centers need to be supplied with sensory information arriving at the trunk, limbs, or tail, which is solved by transmitting it via the spinal cord to the brain. Much of the head-specific, low-level sensorimotor processing is mediated by nuclei in the medulla, pons, and midbrain, collectively forming the larger conglomerate called the *brain stem*. Analogous to the 31 nerves that connect peripheral body parts to the human spinal cord, the brain stem connects to 12 cranial nerves relaying outbound motor information and inbound sensory signals. One of these nerves (the *nervus vagus*) is less concerned with the head and exerts autonomous control over vital organs inside the trunk, including the heart, guts, and lungs.

The *medulla* is organized in a way similar to the spinal cord and exerts control over blood pressure and respiration. In addition, it contains cell groups for control of head muscles, and nuclei relaying gustatory, auditory, and vestibular information to higher brain centers. Relative to the medulla, the pons lies dorsally (i.e., toward the back surface of the

animal; the topographic term “ventral” means toward the belly surface, whereas “rostral” and “caudal” mean “toward the nose” and “toward the tail,” respectively). Similar to the medulla, the *pons* controls vital body functions such as respiration but also contains nuclei controlling sleep and arousal or relaying sensory signals either toward higher centers such as the neocortex or from the cortex down to the cerebellum. Likewise the *midbrain* contains a heterogeneous blend of nuclei—for instance, for auditory processing and low-level control of eye movement. Across the pons and midbrain, we find cell groups that have widespread projections to higher brain centers and synthesize neuromodulatory transmitters. These slow-acting “state setters” include dopamine, serotonin, and norepinephrine and regulate vigilance, attention, emotional state, motivation, and learning processes. Also acetylcholine belongs to this class of central modulatory transmitters and is produced by cell groups in the brain stem’s ascending reticular formation. Acting on thalamocortical circuits, these and other cholinergic cells are particularly important for regulating the sleep-wake cycle and maintaining arousal and consciousness (Steriade, McCormick, et al., 1993; Schiff & Plum, 2000; Steriade, 2004; Crunelli & Hughes, 2010).

The *cerebellum* is a fascinating structure from the viewpoint of representations and consciousness, containing many more neurons than any other brain structure. In humans, the cerebellum contains ~100 billion cells (Andersen et al., 2003), whereas the neocortical count ends at a mere ~20 billion cells (Pakkenberg & Gundersen, 1997). Like the neocortex, it is equipped with pathways relaying sensory inputs and motor outputs. However, in contrast to the neocortex, the cerebellum is not commonly associated with conscious processing. Massive cerebellar lesions cause *ataxia* (poorly coordinated stepping movement, paired with poorly maintained upright posture—“drunken gait”), deficits in learning fine motor skills, and deficits in adaptation of sensorimotor coordination, such as the adjustment of head and eye movements to changing visual input. A recurrent theme is that the cerebellum controls both the *gain* of motor responses to changes in sensory input, such as the strength of an ocular reflex movement in response to head rotation, and the *timing* of a specific motor response relative to ongoing sensory inputs and overall motor patterns (Llinás et al., 1997; Dean et al., 2010; De Zeeuw et al., 2011). Evidence is accumulating for a role of the cerebellum in cognitive processes, such as language and the anticipation of future events.

The thalamus and hypothalamus form the chief components of the *diencephalon*. We already saw how some thalamic nuclei relay sensory information to the neocortex. The term “relay” fails to convey the functional complexity of these nuclei, as it has become clear that they gate and modify the transmission of information toward the neocortex, depending on other inputs they receive, including feedback from the neocortex itself. Cells in the LGN can either assume a depolarized state in which they fire tonically and reliably in response to visual stimuli, or a rhythmic state, characterized by cycles of hyper- and depolarization, in which they fire phasically and intensively in spike bursts, especially during deep sleep (Steriade, McCormick, et al., 1993; Sherman, 2006).

Less well studied are the higher-order thalamic nuclei, which do not receive sensory input via subcortical routes. These nuclei process information from primary sensory neocortical areas and relay this to higher-order cortical areas generally belonging to the same sensory modality. For instance, the pulvinar, located in the caudal thalamus, processes information from the primary visual cortex and innervates higher-order visual areas in the extrastriate cortex, which are associated with analysis of more complex visual features, spatiovisual representations, and object perception. Moreover, the thalamus contains two other groups of nuclei—the intralaminar and midline nuclei—which are considered “non-specific” because they would project more diffusely to cortical and subcortical regions. However, anatomical tracing studies have shown that the main projections of these nuclei, reaching the neocortex and striatum, are more distinct and topologically specific than originally anticipated (Van der Werf et al., 2002). Damage to this group of structures has been correlated to loss of arousal and consciousness and, in extreme cases, to vegetative state or irreversible coma (Llinás & Ribary, 2001; Schiff et al., 2002). Finally, I already mentioned the reticular nucleus, wrapped around the inner parts of the thalamus like an onion’s peel. Its neurons are critical for bringing thalamic relay cells and their cortical counterparts into a state of oscillation, contributing to a regulation of sleep–wake states.

Sitting at the base of the diencephalon, the *hypothalamus* has been classically viewed as part of the “reptile brain,” with its well-known functions in fixed behavioral patterns that are basic to survival and reproduction. Its nuclei and cell groups have been implicated in feeding and keeping energy balance in check, but also in aggression, sexual activity, sleep, and maternal behavior (Sutcliffe & de Lecea, 2002; Blouet & Schwartz, 2010; Woody & Szechtman, 2011; Flanagan-Cato, 2011; Stolzenberg & Numan, 2011). Another function of the hypothalamus is to regulate hormonal secretion, funneled into blood circulation by way of the pituitary gland.

To maintain an optimal balance (homeostasis) in minerals and energy supplies in blood circulation, the hypothalamus and adjacent forebrain nuclei contain *sensory cells*. For instance, cells in the supraoptic nucleus function as osmosensors, gauging salt levels in blood plasma (Bourque, 2008). Neurons in the preoptic area are temperature sensitive and regulate dissipation of body heat (Morrison et al., 2008). Other examples are forebrain sensors of blood pressure, glucose, and nutrient levels (e.g., Kirouac & Pittman, 1999; figure 2.2). Finally, we already met the suprachiasmatic nucleus, harboring our endogenous biological clock. Its pacemaker cells fire more intensely during the subjective day than during the night. Moreover, this clock has a sensory function in that it can be synchronized to the external light–dark cycle via optic afferents (Groos & Mason, 1978; Pennartz, de Jeu, et al., 2002).

2.5 Control Layers in the Brain

Before surveying the telencephalon, let me introduce the concept of *control layers* as a general organizational principle. Going back to *lower vertebrates*, the spinal cord, medulla and pons can be considered the lowest control layer, at which sensory inputs are relatively directly coupled to motor outputs, with the critical addition that descending output from higher control layers can intervene with this coupling. For example, the ocular motor nuclei in the brain stem are controlled by the superior colliculus, which is part of the mesencephalon (Schiller & Tehovnik, 2005; Thier & Möck, 2006). This structure is capable of integrating sensory input from both the visual and auditory modality and converts this joint input into low-level, relatively automatic motor commands for saccadic eye movement (Bell et al., 2005). In turn, however, this system is under control of an even higher-level control layer involving the neocortical frontal eye fields, supplementary eye fields, and posterior parietal cortex in conjunction with inhibitory control via the basal ganglia (Schiller & Tehovnik, 2005). In a similar vein, control over locomotion can be considered to be set across layers, with central pattern generators in the spinal cord being regulated by locomotor regions in the mesodiencephalon. These regions, in turn, are controlled by output from the striatum and pallidal system in concert with neocortical systems for voluntary movement (Garcia-Rill, 1991; Grillner et al., 2005).

Hierarchical processing across serially connected stations can also be recognized in sensory systems. Common to processing touch and pain information, the design is to relay inputs from peripheral receptors via the sensory ganglia and spinal cord to specific brain stem nuclei. Output from these nuclei is then propagated to the mesencephalon and hence to specific thalamic relay nuclei, connecting to primary sensory neocortex (Willis & Westlund, 1997; Willis, 2007). In turn, high-level processing areas such as the cortex provide feedback controlling lower levels of processing. The auditory, gustatory, and vestibular modalities conform to this basic plan; it should be noted that inputs from these modalities are not arriving via the spinal cord but via sensory-specific brain stem nuclei. An exception to this design is the olfactory system, where inputs are relayed via the nasal epithelium and olfactory bulb to the piriform cortex and several forebrain nuclei, to gain direct access to cortical structures such as frontal regions and entorhinal cortex, partially bypassing the thalamus (Gottfried, 2010).

The concept of hierarchically organized control should be applied with caution as it might lure us into an artificial distinction between a lower “reptile brain” and a controlling, repressive neocortical system of Freudian “superego” proportions. A crucial amendment to the concept is that lower-level control layers usually feed back to higher centers. For instance, hypothalamic output pathways reach back to prefrontal cortex, a high-level center for cognitive and behavioral control (Sylvester et al., 2002). In the oculomotor system, the superior colliculus informs the frontal cortex about its ongoing activity (Sommer & Wurtz, 2008). As far as we can say that a “reptile” brain is harbored in the basement of our

cranium, such phylogenetically ancient parts should be considered densely connected and integrated with the higher-level circuitry of the thalamus and neocortex.

2.6 A Bird's Eye View of the Mammalian Brain: Telencephalic Structures

Wrapped up inside the skull like a folded wad of tissue, the neocortex can be divided into four main domains fitting into the corresponding recesses of the skull: the frontal, temporal, occipital, and parietal lobes. Of these, the frontal domain is chiefly concerned with motor functions, which are tightly linked to the primary and secondary motor cortex and the supplementary motor regions. Closer to the forehead and eye sockets, the prefrontal parts of this lobe mediate higher-order, planned control over motor behavior and cognitive processes, such as attention. Frontal regions communicate intensively with the basal ganglia, which are tucked away beneath the neocortex and the thick sheet of fibers connecting the cerebral hemispheres—the corpus callosum (see figure 2.8). The basal ganglia consist of a group of nuclei: the striatum, pallidum, subthalamic nucleus, and substantia nigra—comprising pars reticulata and pars compacta). They process the inputs received from the neocortical–hippocampal cortical mantle, subserving a diversity of functions such as posture and movement control, skill learning, and basic forms of reward-dependent learning and conditioned behavior.

The frontal lobe is anatomically segregated from the other lobes by a deep groove called the central sulcus (see figure 2.9). The two main jobs of the temporal, occipital, and parietal lobes are sensory representation and memory of information about stimuli, events, and facts. This may well be an oversimplification, but here “sensory” is meant to include both low-level, unimodal information and high-level, multisensory integration. For instance, a high-pitch tone activating the primary auditory cortex (A1), located in the temporal lobe, counts as low-level sensory information, whereas the sense of our body position in three-dimensional space, involving the posterior parietal cortex, qualifies as a higher level of representation, as it is dependent on an integration of vestibular, visual, tactile, and proprioceptive¹ inputs.

At the very occipital pole of the cerebrum, we find the primary visual cortex, whereas areas for somatosensory processing are situated in the parietal lobe, just caudal to the central sulcus. Both primary cortices (V1 and SI) project to higher-order regions operating within the same modality, for instance areas V2, V3, and V4 in the visual system, and secondary somatosensory cortex (SII) and Brodmann's area 5 of parietal cortex for tactile sensing. Important properties of a visual cortical neuron are the location and size of its receptive field, which is the region of the visual field to which a neuron can respond when presenting a change in input. At primary cortical levels, cells tend to have smaller receptive fields and respond to simpler stimulus features than cells situated at higher levels. For instance, area V1 harbors simple cells that respond strongly to line or bar elements with a specific orientation in the visual image. Complex cells have larger receptive fields but also

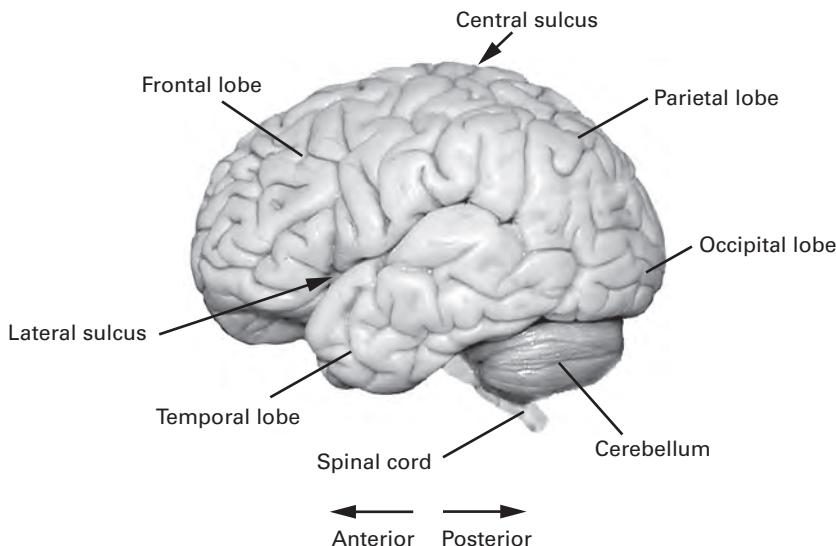


Figure 2.9

Lateral view of the human brain, showing the locations of the major lobes and sulci. Sulci are indicated by arrows. Visual areas are located in and close to the occipital lobe, auditory areas are part of the temporal lobe, and the somatosensory cortex is located just posterior (caudal) to the central sulcus.

respond to oriented bars, albeit their exact position in the visual field matters less. Higher up in the visual hierarchy, the inferotemporal (IT) cortex contains cells with very large receptive fields, responding to specific faces or other complex objects. This hierarchical distinction between lower and higher sensory areas is also found in the temporal lobe, where we find secondary auditory cortex (A2) next to A1. In addition, this lobe houses regions contributing to speech analysis (Wernicke's area) and cognitive processes such as recognition and categorization.

Returning to the parietal lobe, neurons in posterior parietal cortex² not only process low-level somatosensory information but also engage in high-level representations of body and head position and cast this information in a form suitable for conversion into motor acts ("visuomotor transformation"; Goodale et al., 2005; Andersen & Cui, 2009). Whereas the temporal lobe is strongly concerned with the "what" aspects of visual input (e.g., object identity), the parietal cortex engages to a major extent in spatial and action aspects ("where" and "how") and visually guided perceptual decisions.

Given this medley of functions, we are confronted with the question of how all of these cortical areas are ordered. How do they communicate with one another, and is there any hierarchy of functions? David van Essen and his colleagues addressed this question by studying anatomical projections between visual cortical areas, confirming a hierarchical organization. We will pay attention to the notion of "feedforward" and "feedback"

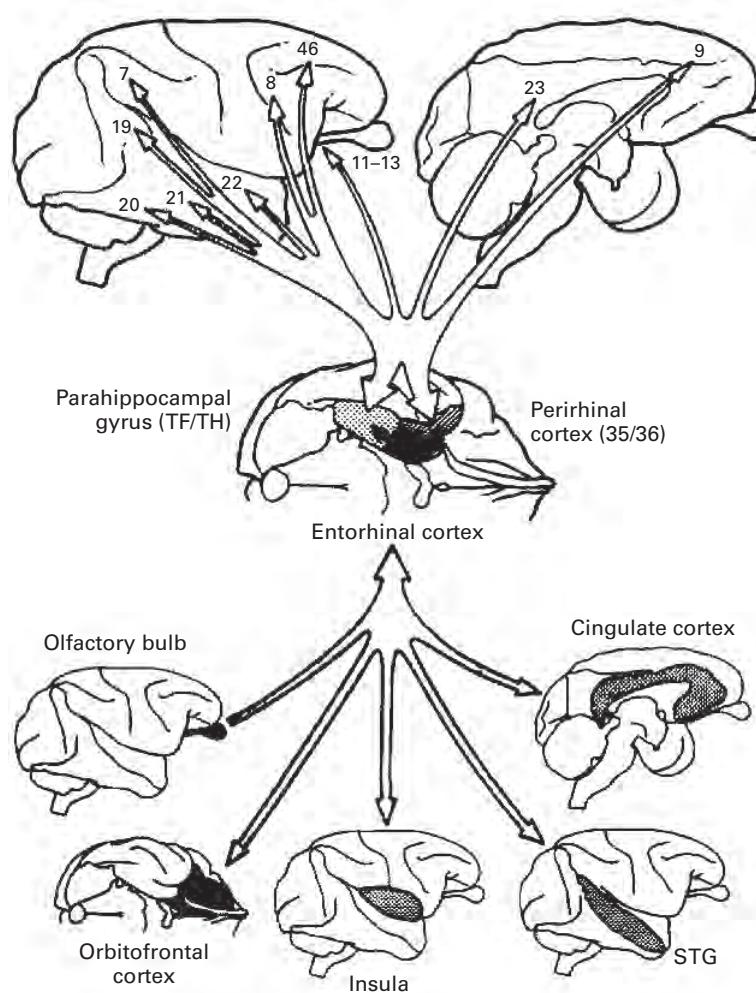
connections between lower and higher neocortical areas in chapter 6. Briefly, their hierarchical scheme (figure 6.3, plate 3; Felleman & Van Essen, 1991) suggests how “lower-order,” primary visual areas project to higher-order areas, which in turn communicate with fewer and fewer structures placed yet higher in the hierarchy, converging to an apex. In turn, these higher areas send feedback (recurrent) projections to their lower sensory counterparts.

From the viewpoint of consciousness and searching for overarching, central representations, the brain area at the apex may be extremely interesting. What is this highly integrative “summit” region, where all inputs come together? It turns out to be the hippocampus,³ placed in a privileged position where it can absorb preprocessed information from all main sensory modalities: vision, audition, touch and pain, proprioception, smell, taste, and the vestibular senses (see figure 2.10; Squire et al., 1989; Amaral & Witter, 1995; Eichenbaum, 2000). This might tempt us to think of this structure as an all-seeing eye, a “seat of the soul.” Yet, this structure does not appear to account for consciousness but forges multimodal inputs into a neural code for *spatiotemporal context*—for multimodal configurations of stimuli in time and space—and has a main function in memory, not conscious perception per se.

This type of code is illustrated by firing patterns of hippocampal cells that change as a function of an animal’s exploration through an environment. Hippocampal cells fire specifically at a location an animal or human occupies in space, largely independent of the behavior the animal is displaying at this location or its head direction (O’Keefe & Dostrovsky, 1971; McNaughton et al., 2006; Ekstrom et al., 2003). This coding of spatial location may enable animals to construct allocentric representations of their environment (i.e., knowledge about the spatial relationships between objects independent of the animal’s current viewpoint; O’Keefe & Nadel, 1978). It is especially enabled by the integration of visual and vestibular inputs with self-motion information, which is derived from proprioceptive feedback as well as information from the motor system. Nowadays, a stronger role is attributed to the entorhinal cortex, adjacent to the hippocampus, when it comes to explaining spatial navigation and self-localization, whereas for the hippocampus itself the emphasis has shifted toward coding and memorizing objects and events in their spatiotemporal context (Derdikman & Moser, 2010; Leutgeb et al., 2005).

Thus, the seduction of neuroanatomy must be resisted when it comes to linking consciousness to any putative “summit region.” In addition to the “upstream” pathways that relay sensory information to the hippocampus, the hippocampal–entorhinal memory system reaches back essentially to the same cortical areas that fed into them, via parallel “downstream” pathways (see figure 2.10; Squire et al., 1989; Amaral & Witter, 1995; Eichenbaum, 2000). This upstream/downstream parallelism may be a key feature for the formation of our long-term memory for facts and personal events.

Despite brave attempts by Popper and Eccles (1977) to link the mind to a singular region in the brain, it now seems very unlikely that such a “seat of the soul” would have

**Figure 2.10**

Many of the “higher” (i.e., nonprimary) neocortical areas are reciprocally connected to the entorhinal–hippocampal system. In the upper part of the figure, bidirectional connections between neocortical areas on the lateral (left) and medial (right) surface are shown, denoted as Brodmann areas. The central panel shows three structures on the ventral side of the temporal lobe: parahippocampal gyrus (areas TF/TH), perirhinal cortex, and entorhinal cortex. These areas process and relay information going into or out of the hippocampus. The lower panel shows additional areas connected to entorhinal cortex, of which the olfactory bulb projects to the entorhinal cortex but is not known to receive a substantial return projection. From Squire et al. (1989), with permission from Elsevier.

any specific location in the brain. One might search for such a location by considering which area is most “central” and connected to most other brain areas. Even more so than applies to hippocampus, the amygdala—which consists of at least 13 nuclei (Pitkänen et al., 1997)—receives inputs from all major, higher neocortical regions (e.g., Stefanacci & Amaral, 2002). Based on anatomical tracing studies, Scannell and Young (1993) showed that the amygdala is placed very centrally in the web of brain connectivity, linked as it is to both cortical regions and subcortical areas such as the striatum, hypothalamus, and brain stem. This strategic position, however, is not utilized for conscious perception but rather for associating emotional value to sensory input and, hence, the adaptive regulation of emotional behavior and autonomous functions such as heartbeat rhythm (Baxter & Murray, 2002; Phelps & LeDoux, 2005).

Two further areas are known for their confluence of widespread afferent input and integrative functions: the cingulate and insular cortices (see figure 2.10). The cingulate cortex (from the Latin *cingulum*, girdle) encircles the corpus callosum at the medial cortical surface. Closely connected with other emotion- and memory-related structures such as the amygdala, hippocampus, and prefrontal cortex, it is implicated in so many functions that its common, overarching function (if any) is still mysterious. The anterior part of the cingulate cortex alone has been implicated in functions ranging from voluntary, spontaneous motor behavior to conscious resting and mind wandering, conflict monitoring, error detection, pain perception, and adaptive selection of effortful behavior (Gehring & Fencsik, 2001; Greicius et al., 2003; Craig, 2002; Yeung et al., 2004; Dalley, Cardinal, & Robbins, 2004; Walton et al., 2007). Although more research is needed to determine a common function underlying this diversity, the weight of current evidence supports a role in the coding of, and responding to, discrepancies between expectations and actual outcomes of situations and actions.

Hidden beneath the lateral sulcus that separates the temporal from the frontal and parietal lobes, the insular cortex (or “insula”) receives a multimodal flux of information on the state of the body. Similar to the amygdala and cingulate cortex, the insula is important for the affective processing of input, ranging from nociceptive body information to sexual excitation and orgasm. Insular lesions do not abolish pain perception per se, but flatten emotional reactivity and the subjective experiencing of pain. The insula also integrates other types of information besides pain from sources close to, or inside, the body, including taste, visceral sensations, and touch. Craig (2002) reviewed spinothalamic pathways projecting to the insula that are activated when the skin is gently touched or stroked, representing an emotional modality of “pleasant skin feeling.”

2.7 Maps: The Tessellation of Sensory and Motor Space

When discussing neural representations in the context of conscious or unconscious perception, sensory maps in cortical or subcortical structures present a key issue. But first we

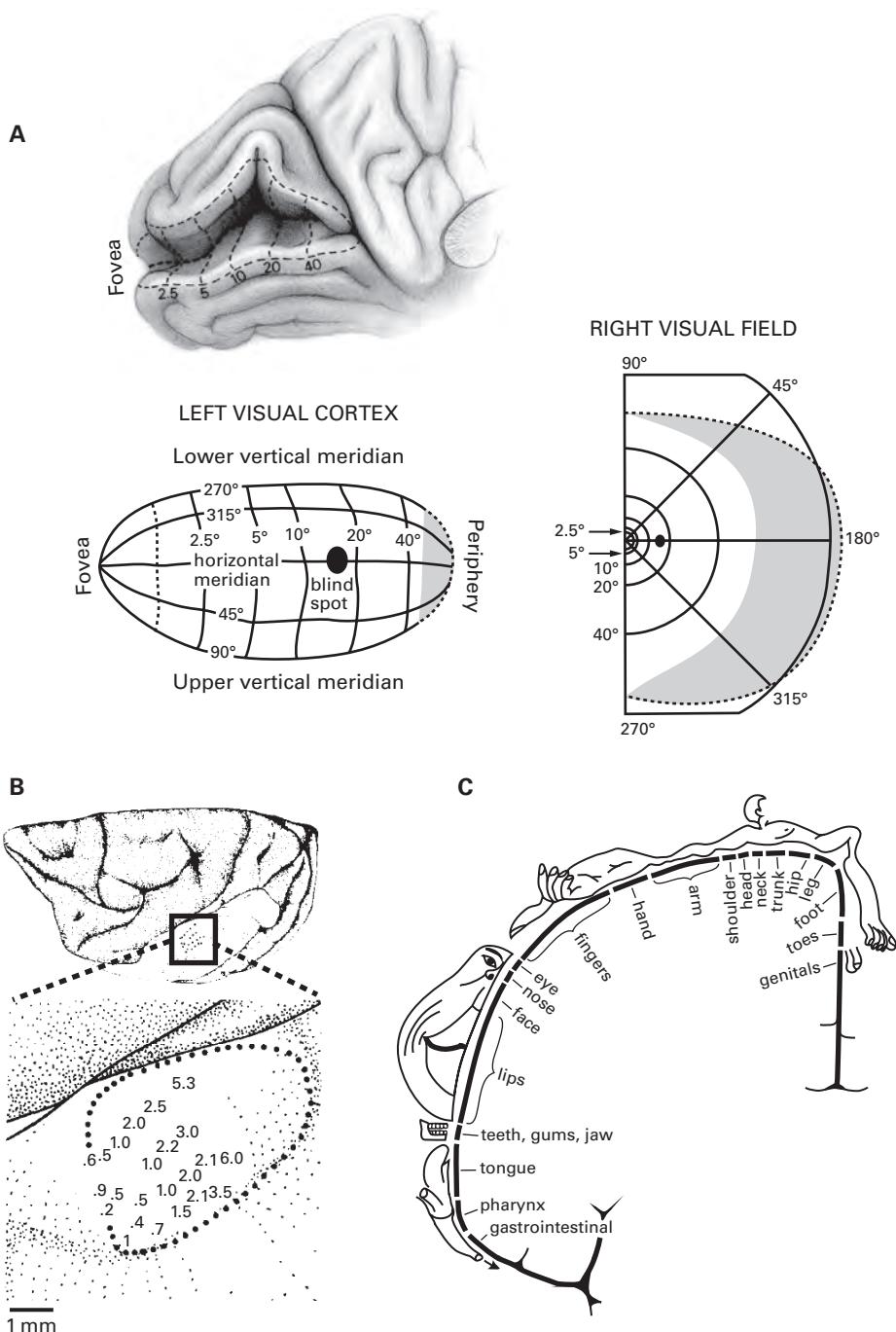
must ask this question: What qualifies as a map? A conventional geographic map is a two-dimensional diagram of a segment of the earth's surface showing features such as land, sea, cities, and mountain ranges. Which features are represented by symbols on the map may vary widely, but for the map to be functional it is essential that human observers can translate these symbols into concepts they understand, such as a little red square meaning "state capital." A second critical feature is a topographical arrangement: neighboring cities are represented as being close together on the map, remote cities are far apart, and relative spatial relationships between groups of objects are preserved. An analogue or digital camera picture qualifies as a map of a visual scene we have seen, notwithstanding that the picture lacks obvious symbols that we would need to interpret first.

Throughout the brain, many types of maps are found, both at subcortical and cortical levels. Some maps have a layout based on external space whereas others are based on a different coordinate system. Most well-known is the retinotopic mapping in V1, but it is also found in at least 20 extrastriate visual areas such as V4 and V5/middle temporal (MT) in primates (Allman & Kaas, 1971; Maunsell & Newsome, 1987; Felleman & Van Essen, 1991; McKeefry & Zeki, 1997; Tolias et al., 2001). Light patterns impinging on the eye are first inverted via the lens and projected onto the retina, where photoreceptors transduce photon inputs into electrical output, which in turn is transmitted to the retinal ganglion cells via complex preprocessing steps. Spike patterns of these cells are transmitted via the optic nerves and LGN to the primary visual cortex. Retinal topography is preserved in V1 even though the center of our visual field, projected onto the fovea of the retina, is massively overrepresented here (see figure 2.11A). The central portion of our visual field, captured within a circle spanning no more than 2.5° around the fixation point of our eyes, demands

Figure 2.11

(A) Upper panel shows the location of the retinotopic map in the primary visual cortex of the human brain. Groove marked by bold dotted line is the calcarine fissure, which is shown opened for illustration purposes. The fovea is the spot on the retina corresponding to the center of gaze. The relative distance to the fovea is plotted in degrees. The lower left panel shows the spatial layout of the retinotopic map projecting the right visual field onto the left striate cortex, with the peripheral visual field represented on the right (gray). The lower right panel shows corresponding angular contours in the visual field, with the gray zone representing the monocular temporal crescent, where visual stimuli are visible to the right eye only. The horizontal meridian connects the center of the visual field to its extreme periphery in the horizontal visual plane (zero degrees; this line follows base of fissure in upper panel). The lower half of the visual field, bound to the lower vertical meridian, is projected onto the dorsal half of the retinotopic map, and vice versa for the upper half of the visual field. Adapted from Horton and Hoyt (1991).

(B) Upper panel shows location of the auditory tonotopic map in the superior part of the temporal lobe of the macaque brain. Inset shows the tonotopic map in more detail. Numbers are the frequency (in kilohertz) of tones to which neurons recorded at the respective locations respond most. Note that the tonotopic map in auditory cortex is characterized by bands or strips responding to similar frequencies. In the figure, these bands roughly run from top left to lower right. From Merzenich and Brugge (1973), with permission from Elsevier. (C) Body map in the human somatosensory cortex. Visceral–oral areas are represented on the ventrolateral surface, whereas facial, hand, trunk, and limb areas are situated at progressively more dorsal and medial locations. Adapted from Penfield and Rasmussen (1950).



a huge cortical territory relative to more peripheral parts. This implies a strong distortion of world-to-brain mapping, but nonetheless the essential map characteristic of a spatial topography is present.

The tonotopic map of primary auditory cortex presents a stronger deviation from a classic road map: neighboring locations in brain space do not represent sound sources close to each other in external-world space, but inputs of similar pitch or tone frequency (see figure 2.11B; Merzenich & Brugge, 1973). This layout by frequency goes back to the tonotopic map already found along the basilar membrane of the cochlea, where the lowest frequencies we can hear (~20 Hz) excite hair cells at the apex, whereas the cochlear base responds to the highest frequencies (~20 kHz).

In the literature, conscious perception is often exclusively associated with thalamocortical processing, and thus with cortical maps, but why would this exclusivity hold? Many types of map are already found at subcortical levels of processing, not commonly associated with perception. For instance, the midbrain's inferior colliculus contains a tonotopic map (Malmierca et al., 2008), and it is as yet unclear why this subcortical map would not be associated with perception, in contrast to its cortical counterpart. In chapters 6 and 7 we will revisit this issue.

This question also takes us back to the use of symbols or signals indicating features on a map. Who or what reads all of these brain maps? This question is meant to be provocative: the idea of a homunculus in the brain, acting as map reader or internal observer, would lead us back to an irresolvable Cartesian dualism. However, if there is no central agent, are there perhaps other brain areas which collectively “read out” the signals on the map? Could we not say that the posterior parietal cortex reads out information from visual cortical areas? Surely one can argue that activity patterns in sensory maps are transmitted to target areas in “higher” brain systems, which convert these sensory inputs into outputs influencing cognitive processing and behavioral decisions. Especially studies on coding of information by groups of neurons have popularized the notion of target areas “reading out” afferent input (Deneve et al., 1999; Van Duuren et al., 2008). My argument here is to be very cautious in using this term when it comes to perception. It raises the expectation that the target area has some interpretive capacity, being able to “make sense” of lower-level sensory input and attributing meaning to it. This is usually far from empirically established. Talking about a readout in higher brain areas essentially defers the problem of meaning.

A further assumption on sensory maps is that topographically organized neural coding of information is fixed over time. On an ordinary road map, “blue” means “water” and “yellow” means “land,” and this color coding remains set no matter who reads the map or how we use it. But does this assumption also hold for the brain? Does neural activity in a hot spot on the map in V1 represent exactly the same information when a sensory input freshly arrives from the thalamus, as when this hot-spot activity has affected higher cortical areas which feed back to V1? Is the sensory representation in V1 still the same after

receiving feedback, or could it change in form or content over time? This is another dogged issue that will cling to discussions on maps further on.

The diversity of mapping systems can be appreciated by looking beyond the visual modality. The primary somatosensory cortex SI contains four areas (Brodmann's area 3a, 3b, 1, and 2), situated along the gyrus that lies just posterior to the central sulcus (see figures 2.9 and 2.11C). Each of these four regions contains a complete body map, which is—ironically in the context of consciousness—depicted as a distorted homunculus stretched out along the sulcus (Darian-Smith, 1984; Hofbauer et al., 2001). Additional body representations are found in the secondary somatosensory cortex (SII), which processes information received from SI, and in the insular and posterior parietal cortex, to which SII projects.

A given modality can be represented by multiple kinds of maps that each provide information about different attributes of a stimulus. Sound, for instance, is more than just frequency or pitch, composed as it may be of speech or a melody with a rich rhythmic and harmonic structure. Such higher-order aspects of sound structure are thought to be represented in cortical auditory regions receiving inputs from A1, primarily lying in or close to the temporal lobe, and traditionally associated with Wernicke's area (Narain et al., 2003; Zatorre et al., 2007). Conversely, multiple sensory modalities can contribute to a single map. Imagine you are walking through a forest and a bumblebee flies straight toward your face. The human parietal face area, receiving convergent tactile and visual information, is likely part of a rapid response system that maps multimodal inputs in peripheral space and helps keep the looming insect at bay (Sereno & Huang, 2006). Subcortically, a similar confluence of tactile, auditory, and visual input is found in the superior colliculus, which commands orientation responses toward sources in either modality (Stein & Meredith, 1993; Stein et al., 2001). Perhaps the ultimate example of a multimodal map is provided by the hippocampal formation, which, as we have seen, can represent an animal's spatial position because of a confluence of information from the visual, vestibular, kinesthetic, and other senses.

2.8 A Multitude of Maps, yet One Common Viewpoint

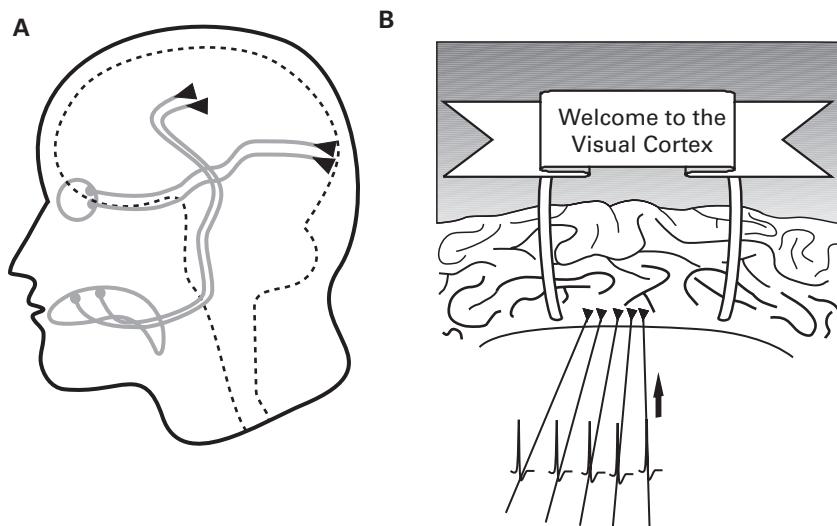
The ubiquitous occurrence of maps throughout the brain confronts us with deeper problems, especially if we wish to explain subjective experiences in neural terms. Given that the brain contains very different maps, how is it that we subjectively experience the world, including our own body, as a unified whole, and perceive it from a single perspective—our own current viewpoint? This first-person viewpoint has been emphasized as a hallmark of conscious experience by many philosophers (e.g., Searle, 1992, 2000; Churchland, 1995; Metzinger, 2000). Sensory maps differ not only in the modality in which they are set but also in their frame of reference. That V1 has a retinotopic frame of reference, for instance, implies that every eye movement will cause a spatial shift in the activation pattern of V1's retinotopic map, as the location of the receptive fields of V1 neurons is defined relative to the fovea. These neurons do not correct for eye movement. Yet, as we move around our

eyes to explore a wall-size painting, we experience the painting not as jumping around in our field of view but as being static relative to our own body. This visual stationarity may involve neurons in posterior parietal cortex, which do appear to correct for eye movements (Zipser & Andersen, 1988; Galletti et al., 1993). They are thought to obey a *craniotopic* (head-centered) frame of reference because their response to a visual object varies with head rotation. Yet higher-order sensory maps exist, taking into account both eye and head movements to yield a body-centered (or “egocentric”) reference frame. These maps may be associated with the dorsal parietal–frontal stream of sensory processing (Goodale & Milner, 1992; Committeri et al., 2004; Neggers et al., 2006). Which of these reference frames corresponds most closely to the first-person perspective we experience in daily life? This question will be addressed in chapters 8 and 9, where I will argue that “having a perspective” is not so simple or unitary as it appears. As our eyes scan our environment and our head rotates to support this exploration, our view on the environment changes, yet we can also sense, more in the background, that our limbs and body either remain where they are or move around in space (Galati et al., 2010).

2.9 Further Puzzles on Sensory Maps: What Makes Visual Cortex “Visual”?

The problem of first-person perspective already presents a major challenge, but an even deeper problem is how sensory systems are able in the first place to generate activity patterns to which our brain attributes content, or phenomenal meaning. Suppose we build a mock imitation of the visual system by connecting a video camera to a two-dimensional array of LED lights that will flicker on and off as they receive signals corresponding to pixels in the camera’s chip going black or white. It would not be meaningful to say that the LED array generates or has some kind of visual experience of the input from the camera—but we did actually construct a kind of sensory map!

So what makes neural activity in the primary visual cortex V1 “visual”? Is it enough to say that the neurons *respond* to visual input? The problem here is that neither these neurons, nor their fellow circuits in higher areas, have any *intrinsic* knowledge specifying the kind of information they are processing. Their sensory modality is not a built-in property of the cortical area they are located in. Will it have to be *afferent input* that determines the content or meaning of what their firing activity represents? In primates, area V1 is in the unique position that it is virtually the only cortical area receiving direct and strong input from the visual thalamic nucleus (Kennedy & Bullier, 1985), so this unique input relationship can be argued to be sufficient for specifying the modality that V1 will process. But the problem here is that no information is present, at the level of V1 neurons, *where their input is coming from*. As beholders, we can observe the anatomy of the visual system from an outside viewpoint and trace back optic inputs to the retina. However, the visual system does not possess intrinsic information about the anatomy of the afferent inputs reaching it. When action potentials arrive in V1 along axons from the LGN, there is no “label” or

**Figure 2.12**

(A) Via several relay stations (not shown), taste receptors in the tongue project to the gustatory areas of the human neocortex, mainly located in the insula. Photoreceptors in the retina similarly project to the visual cortex. This external perspective on brain anatomy indicates how different brain areas are selectively provided with information from different peripheral sensors. At the level of cortical neurons, however, no information is available about the sensory origin of their inputs. (B) When spike trains arrive along axons in the visual cortex, they come without “label” or “address of sender,” and there is no “welcome” sign indicating that the spikes have entered the visual cortex, as opposed to cortical areas coding other modalities. This raises the question of how the brain itself identifies the modality to which incoming spike trains belong.

sign that says “Welcome to the Visual Cortex” (see figure 2.12). As far as a V1 neuron is concerned, it just receives synaptic inputs, and these are transmitted all the same across cortical areas, no matter to which modality they belong.

This point is reinforced by cortical anatomy and the brain’s ontogeny. The cortex is characterized by *self-similarity*, implying that the basic cell types and wiring are very similar across different areas involved in very different functions. This is a simplifying statement, because some cortical areas exhibit nontypical features in cytoarchitecture, such as the lack of a developed L4 in regions of the prefrontal cortex. Yet, these deviations can be considered variations on the overall theme of canonical cortical organization (Douglas & Martin, 2004). The cortical mantle thus finds itself in the curious situation that its regions are involved in very different functions—for example, visual perception, smell, language, geometry, emotions, and motor planning—yet the basic circuitry mediating these functions is more or less the same.

Studies on brain ontogeny have produced fascinating insights into the development of modality-specific processing. The group of Migranka Sur showed that the mammalian neocortex is ontogenetically “equipotential” in the sense that a given cortical area

(e.g., auditory cortex) that is anatomically predisposed to mediate auditory processing can come to process information from the visual modality (Roe et al., 1990; see also Metin & Frost, 1989). This was demonstrated in young ferrets by rerouting axonal pathways relaying visual information to the medial geniculate nucleus, so that spike patterns from the “alien” (visual) modality were fed into the thalamocortical circuitry normally dedicated to auditory processing. The “auditory” cortex then developed a spatial map showing typical visual features such as orientation selectivity, and ferrets utilized this rewired map in a behaviorally meaningful way, enabling the animal to make visually guided decisions (Von Melchner et al., 2000). A similar argument has been presented for visual cortex tissue, which was experimentally reorganized to process tactile information when grafted into the somatosensory cortex (Schlaggar & O’Leary, 1991). In summary, cortical tissue is not “predestined” or genetically preprogrammed to process a particular sensory modality but becomes dedicated by the input patterns it receives during brain development. The problem has not been settled and will be further explored in chapters 8 and 9. For now, we conclude that the question “What makes neural activity in cortical area V1 visual?” has no easy, natural answer.

2.10 From Maps to Representations and Projections

What do we actually take the concept of representation to mean? In its most general sense, a representation is an entity that conveys information about something else by being similar to it in one or more of its properties. For instance, an oil painting of Winston Churchill tells something about Winston Churchill when it resembles him in the visual layout and colors of his facial features, but not through the substance it is made of (oil paint vs. flesh, skin, and bone). Representations—as “copies” of an original—can take many forms, including visual images, statues, and models of objects and persons, but can also be abstract and symbolic renderings of a physical regularity, such as “ $E = \frac{1}{2} mv^2$ ” as a representation of kinetic energy. Personal computers use an internal digital representation of color, such as the hexadecimal code “0000FF” for “blue.”

This broad concept of representation might be fine as an initial framework for thinking about sensory representations in the brain. We can say that retinotopic maps in visual cortical areas generate visual representations, even though it remains unclear how the activity of neurons does something more than a television does when representing a part of the external world on its LCD screen. Broad-sense “representation” should not lure us into believing that its widespread application to sensory systems explains how conscious representation arises. Conscious representations present a radically different subspecies of representation in that oil paintings have no experience of what they represent—at least they do not present any evidence that they would. As I will argue, it is unlikely that an isolated patch of electrically active visual cortex would generate conscious representations, but we can say that the brain as a whole does this. Throughout this book I will use “representation”

in its broad sense, but it goes accompanied by the caveat that the term itself has virtually no explanatory power on consciousness.

For some readers, this warning may be completely superfluous, but the following quote from Kandel et al.'s (1991) third edition of *Principles of Neural Science* suggests the distinction should be explicitly made:

Sight, sound, touch, pain, smell, taste and the sensation of bodily movements all originate in sensory systems. These perceptions, in turn, form the basis of our knowledge about the world. Perception begins in receptor cells that are sensitive to one or another kind of stimuli. Most sensory inputs are perceived as a sensation identified with a stimulus. Thus, short wavelength light falling on the eye is perceived as blue, and sugar on the tongue is perceived as sweet. (p. 327)

This quote raises deeply mysterious notions. How is it that perception—understood as “conscious sensation”—could begin in receptor cells, where only a physical transduction takes place from one kind of energy (e.g., air pressure waves) into another (electric activity)? If “sensory input” can be equated with electric activity of receptor cells, why would this result in perception and not in a nonexperiential representation such as in a digital photo camera? Who or what takes care that the inputs are identified with (or as) a stimulus? We have a long way to go when explaining how short wavelength light falling on the eye may correspond to a perception of “blue.”

There are further reasons to be careful about “representations.” From its Latin origin, a representation is meant to “re-present” something, or present an object *again* in addition to having the object somewhere in its original form. A representation of pink would be a re-production, re-creation, or re-construction of the original color pink present in external space. Given the physical notion of electromagnetic waves of specific frequencies hitting our retina, this assumption is difficult to justify without further philosophical contemplation. Our brains are set up to consciously represent the color pink as belonging to an object in our external environment—which makes perfect sense if we use our other modalities and our motor system to explore this space or use a spectral analyzer to gauge the wavelength composition of the object’s reflected light. However, it is problematic to claim that this conscious experience somehow represents a “copy” of the external world’s “original.” It may be wiser to adopt an agnostic stance on this question, at least for the moment, and say that the brain *presents* pink, regardless of whether this is a reconstruction of something very similar “out there” or actually a *de novo* construction based on an input derived from something that could be of a different nature than what we perceive (see chapter 11). It is not trivial to infer, over and above our senses, what exactly goes on in external reality.

Part of the confusion about (re)presentations has to do with localization: how could we assign a representation to a position inside or outside the brain? For instance, in his entertaining book *Proust Was a Neuroscientist*, Jonah Lehrer (2007) writes the following:

If sight were simply the retina’s photoreceptors, then Cézanne’s canvases would be nothing but masses of indistinct color. His Provençal landscapes would consist of meaningless alternations of

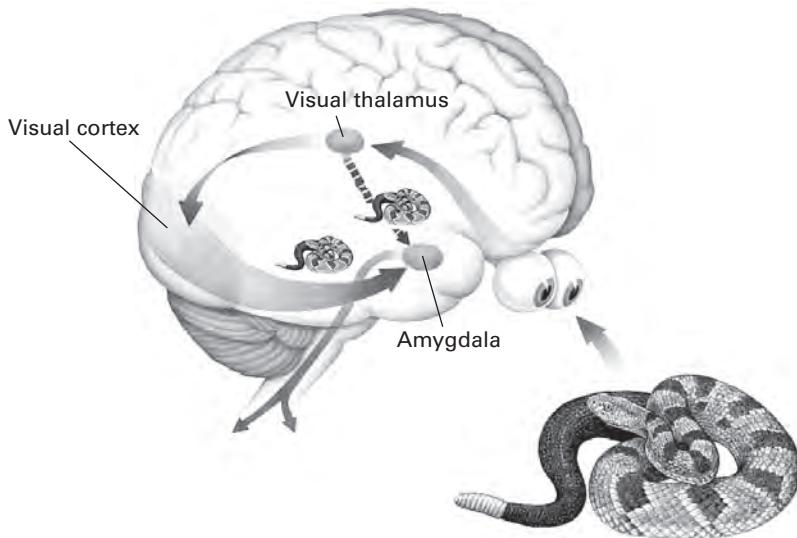


Figure 2.13

Do images of snakes travel along the optic tract and through the brain? When external objects are projected onto the retina, its ganglion cells generate spike patterns that are propagated along the optic tract to the lateral geniculate nucleus (a visual thalamic nucleus) and hence on to the visual cortex, amygdala, and other structures. These spike patterns are associated with the visual input, but it is a different matter to claim that therefore the image of the snake is also propagated along nervous pathways. It is difficult (if not impossible) to point to a location in the brain where the snake image would be. Adapted from LeDoux (1994). It should be noted that the author used the original illustrated for different purposes than intended here, namely, to illustrate pathways in the brain involved in emotional memory and fear responses.

olive and ocher, and his still lifes would be all paint and no fruit. Our world would be formless. Instead, in our evolved system, the eye-ball's map of light is transformed again and again until, milliseconds later, the canvas's description enters our consciousness. Amid the swirl of color, we see the apple. (pp. 104–105)

One of the powerful intuitive (but arguably erroneous) notions exposed here is that spike patterns, traveling across axons or arising in a sheet of postsynaptic neurons, represent some kind of “description” or “image” that is transported from one brain region to the next. This position would force us to commit ourselves to a framework where image-like representations are first processed unconsciously (beginning in the retina and across the optic tracts) and then, by an unresolvably mysterious act, suddenly “reach” the stage of consciousness when arriving at a higher stage of processing. Can spike patterns, arising from photic inputs to the retina, be conceived of as unconsciously processed “images”? Images are a convenient, shorthand way to depict what a generic representation could be, but it is problematic to imagine what an image-in-the-brain would be like in unconscious form (see figure 2.13). Perceiving an “image” implies not only a certain information content but also

an entity that can be viewed and interpreted. In the case of spike patterns propagated across the optic tract, it would be hard to maintain that these patterns are viewed or interpreted at that stage of processing.

Here we are also facing a dichotomy in what “image” means. If by “image” we mean a view or photograph of a snake, the image should be taken as being part of the context in which the snake is seen: “I saw the snake on the path right in front of me,” or “After I ran back a few yards, I took a photo of the snake which I am now holding in my hand.” In this case the snake is assigned to the larger spatial, external context or scene it is viewed in. However, if “image” denotes an active neural process or state by which the brain represents objects (cf. Damasio, 2000), we will have to search for a location of this representational neural process. The difference lies in the idea that brain representations are of a constructive (or projectional) nature. The brain generates a representation of a snake, but *what* is represented is not constructed as part of our neuroanatomy, nor is it localized to it, but projected into an external spatial context. This is also illustrated by phantom pain. A limb amputation may result in feeling pain in an arm no longer present, as the nociceptive circuits in the somatosensory cortical circuits are activated, either spontaneously or by some triggering input from another brain area. The pain sensation is generated by circuits involving the “arm” section of the homunculoid body map (see figure 2.11C), but that sensation is not *felt inside* the brain (or localized in it), but projected onto a body part no longer physically connected. The way our senses conspire to construct this externally projective capacity will be examined in chapters 8 and 9.

3

Peeling Off Our Conscious Lives

The “sensible quality,” the spatial limits set to the percept, and even the presence or absence of a perception, are not *de facto* effects of the situation outside the organism, but represent the way in which it meets stimulation and is related to it. [...] The function of the organism in receiving stimuli is, so to speak, to “conceive” a certain form of excitation. The “psychophysical event” is therefore no longer of the type of “worldly causality,” the brain becomes the seat of a process of “patterning” which intervenes even before the cortical stage [...] The excitation is seized and reorganized by transversal functions which make it *resemble* the perception which it is *about to arouse*.

—M. Merleau-Ponty (1945, pp. 86–87)

3.1 What Makes Up a Conscious Experience?

Many countries around the world are witnessing a steady, sometimes soaring rise in the public interest in psychology and the brain. Given this trend, it would not be too surprising if an extremely interested and thrill-seeking volunteer, whom I will call Harry for convenience, would step up to undergo some radical experiments. The aim of the experiments would be to dissect conscious processing, by reversibly inactivating brain structures mediating different functions, and then study whether Harry’s consciousness would stay intact, be lost, or be modified. Considering their medical risks, such experiments would not normally pass prevailing legal and ethical standards, but for the sake of our thought experiment we assume they would.

Technically speaking, the experiments are feasible in principle, as a wealth of methods is available to inactivate—or boost—activity in specific brain areas. These techniques include, for instance, deep brain stimulation, by which neural circuits can be inactivated for prolonged periods of times—for instance, to suppress Parkinsonian tremor. Another method that is well established in animal experiments is the injection of a neurochemical substance into a specific brain area that will temporarily and reversibly block local activity, which is important for understanding the area’s function in behavior. With the advent of optogenetics, it has become possible to inactivate a brain area, or a specific cell type, by light pulses (Chow et al., 2010; Yizhar et al., 2011). In the thought experiment Harry wishes to be conducted on his own brain, we will draw on some of the results gained with

these techniques in humans or animals. Equally informative will be the neurological literature on patients who were unfortunate enough to suffer brain damage by stroke, neurosurgery, accidents, or casualties of war. Before embarking on our thought experiment, let us first ask which cognitive functions we normally associate with consciousness and which properties we can ascribe to it.

Before undergoing the experiment, we ask Harry to describe his “conscious world.” This will give us a sense of what *he* feels conscious experience is about and which of its elements may be targeted in his inactivation experiment:

Well—where would you like me to start? This is about everything I come across as long as I am awake—everything I see, hear, smell, touch, feel under my skin. Everything I experience is played out against some basic emotional feeling or mood. Seeing an autumn forest with all its red and yellow leaves makes me feel calm and joyful. It also makes me want to walk in there. When I go in there, it’s wonderful experiencing this ongoing flow of colored trees coming at me and passing by. I can pick out a tree that strikes me in particular or experience the whole scene at once. I can close my eyes and imagine the forest—it gives a special feeling that it is *me* who is doing all this and can talk about it. That’s it in a nutshell, and it all goes on until I go to sleep at night.

3.2 Sensory Modalities

Many aspects could be added to Harry’s description, but some key elements stand out. *Perception in individual sensory modalities* is central to our daily mental life. In addition to vision, hearing, touch, olfaction, and taste, our multisensory palette comprises pain, thermoception, proprioception, and the vestibular senses. Due to a recent revival of interest in the visceral senses (Damasio, 2000), they have come into the spotlight of consciousness research, bringing up the question of whether they too should be regarded as a separate modality. For the purposes of our inventory, the visceral senses will be embedded in the wider concept of “interoception,” referring to sensations from within the entire body, not only from the gastrointestinal system but including affectively charged sensations such as pain and itch (Schnitzler & Ploner, 2000; Olausson et al., 2002; Craig, 2002). Altogether, the existence of many distinct sensory modalities makes our conscious life fundamentally “rich”—with multiple dimensions that bring about an incredible variation of individual sensations. This variation is distinct from the variance we can measure or experience along a single dimension, such as the fluctuating intensity of the smell of French toast we may experience when sleeping late.

The existence of sensory modalities enriches our life with qualitatively distinct experiences. Vision is radically different from hearing, and touch has a completely different “feel” to it than taste. Even within the domain of one (main) modality we can distinguish different qualitative properties. In object vision we recognize properties such as shape, movement, color, depth, texture, and transparency; I will refer to such attributes as submodalities of vision.

Viewed from classical philosophy, the modalities suddenly put us on the battlefield of the “qualia.” Roughly speaking, the concept of qualia refers to the same qualitative properties of experience that we associate with sensory modalities and submodalities. Qualitative aspects are thought to evade quantification by physical measurement, as they are fundamentally “subjective”: what we smell when inhaling the moist vapor of a hot, Irish stew, and what we feel when our elbow is intensely itching. Qualia go a long way back in history. John Locke (1667) was the first to put them clearly on the map of empirical-philosophical debate as he distinguished two different kinds of object properties. Primary properties, he argued, are object features that can be measured and quantified, such as length, depth, shape, and motion. Secondary properties, on the other hand, are features lacking an immediate quantification or mechanistic account, and these would include color, odor, heat, and sound (cf. Shoemaker, 1990). Across centuries, our progressive understanding of sensory and perceptual mechanisms has largely eroded the distinction made by Locke. Nevertheless, in contemporary philosophy the qualitatively differentiated nature of experience has still been argued to be the hardest problem in consciousness research (Chalmers, 1995, 1996). As such the problem contrasts with “easier” problems such as how memories, emotions, planning, and willed actions contribute to mental states (not that these functions are so easy to understand; Chalmers’s point is that they are more amenable to experimental investigation by current methods).

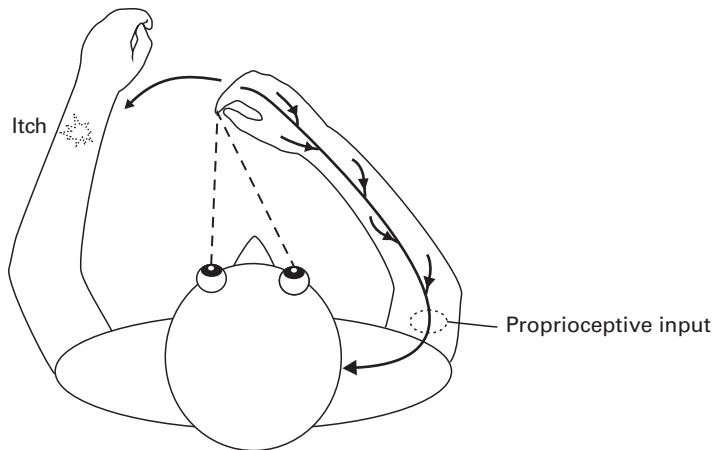
Some philosophers have argued that “qualia” do not really pose a problem that would be harder or easier than related questions on consciousness (Dennett, 1991) or that the problem is real but unanswerable (McGinn, 1989). I will argue, contrary to Dennett, that qualitative properties are an inalienable aspect of phenomenological experience and do pose a problem that demands a scientific explanation. The problem of qualitative, multimodal experience is *not* addressed by asking, “Do you hear this melody the same way as I do?” This question would bring us into the swamp of interpersonal verification of experiential content and private-language problems (cf. Wittgenstein, 1953), but here we stay at a distance from that particular problem. Instead, we ask whether you would describe your *own* experiences as qualitatively differentiated, with the smell of cow dung radically differing from Beethoven’s Seventh Symphony, second movement—regardless of how your neighbor feels about this. This is a within-subject check that every reader may perform for himself or herself. I will assume the test turns out positive.

Whereas perception has often been studied within the setting of an individual sensory modality, Harry said he can “experience the whole scene at once.” This is a complicated point but does align with our common experience that what we perceive appears *integrated*, unified, or “nonfragmented” and, moreover, is ordered in a *framework of space and time*. This framework seems centered on our own body position, follows the orientation of our head and eyes as they peek into the world, referring to the notion of *first-person perspective* encountered in chapter 2. But what do we mean by “integrated”? Sensory integration arises in multiple ways and across multiple dimensions. For example, our two

eyes receive two different visual inputs, yet we perceive only one scene situated in front of us. Our brains utilize this difference (visual disparity) to estimate how deep an object is positioned away from us in the scene. We do not perceive two different images of this scene—they are merged, or integrated, into a unified, three-dimensional view of the world.

When we scan a picture or scene by rapid eye movements called “saccades,” the retinal image (i.e., the spatial pattern of photons impinging on the retina) is changing with every eye movement. Despite the fact that the information of this “jumping” retinal image is transmitted via the thalamus to the neocortex, we do not perceive a jerky, perpetually shifting image of the world. One mechanism to prevent such a dazzling representation from arising is saccadic suppression: the phenomenon that information transmitted during eye movement is not processed to the extent it can contribute to perception (Wurtz, 2008). But even with saccadic suppression in place, our brains have to take into account that an image x acquired at time t is different from, yet continuous with the image y at the next moment $t + 1$, with the eye moving in between t and $t + 1$. How this correction is accomplished is not well understood, but the problem provides another example of why integration is needed to achieve stable perception.

Besides these two examples, our brains have other ways of rendering our experiences integrated and smoothly connected, across both time and space. Different properties or attributes of an object are processed predominantly by different, sometimes remote brain structures. For instance, the motion and color of a single object are thought to be analyzed in distinct higher regions of the visual system (cortical areas MT/V5 for motion and, among others, areas V4/V4 α in humans; Zeki, 2001). Yet, our perceptual machinery flawlessly integrates these attributes into one and the same object. This type of integration amounts to solving the *binding problem* (cf. Milner, 1974; Von der Malsburg, 1981, 1995, 1999). This problem can also be illustrated by two objects being present within a single visual scene, such as a white cup and a silver spoon next to it. Assuming the shape features of the cup and spoon are both processed and represented in particular inferior temporal areas (e.g., Yamane et al., 2008), but their color information is represented in area V4/V4 α and other higher visual areas, how can the brain figure out that the white color does not belong to the spoon and the silvery color not to the cup? Possible solutions to the binding problem will be addressed in chapters 7 and 9. Part of the search for a solution is to examine whether the assumption of segregated processing is actually correct (cf. Edwards et al., 2003). Importantly, the problem of binding—as a specific form of perceptual integration—extends across sensory modalities. When you feel an itch on your left hand and reach out with your right arm to scratch it, you see and feel your arm as it is moving. Information from the visual and proprioceptive senses is seamlessly integrated. But how does the brain “know” that the arm’s movement as observed visually pertains to the same movement as detected kinesthetically? In theory, the kinesthetic information might as well pertain to the left arm or the calf muscle; therefore, mechanisms for multimodal integration will be needed (see figure 3.1).

**Figure 3.1**

An example of multimodal integration. A person feeling an itch on his left arm moves his right arm over to scratch while observing the movement visually. Kinesthetic information is simultaneously transmitted to the somatosensory cortex via proprioceptive inputs (only drawn here for the right, moving arm). Proprioceptive information from both arms is represented on the body maps of the somatosensory cortices. There is no a priori knowledge in this system that the visual motion input should be integrated with the proprioceptive input from the right but not left arm. Therefore a mechanism for multimodal integration is needed.

3.3 Continuity and Working Memory

Much earlier than Harry, William James (1890) already observed that consciousness is experienced in a continuous flow or “stream.” But how does the brain endow us with a sense of temporal continuity? This question should be considered across different time scales. For fine temporal resolution, phenomena such as saccadic suppression and eye blinking indicate that visual input can be briefly interrupted whereas we are not aware of such gaps in sensory supply. However, does this imply there should be a mechanism actively at work in the brain to “glue” the periods of sensory input together? Dennett and Kinsbourne (1992) noted that the “representation of temporal properties” in a mind–brain system is not the same as “the temporal properties of representations.” In a similar vein, we may add that “the absence of representation” is not the same as “the representation of absence.” The lack of sensory input during an eye blink need not be actively represented in the visual system. In fact, the lack of such a representation may be the very cause of not noticing this gap, thus accounting for a feeling (or illusion) of continuity. The same principle may apply when we have the feeling of grasping a whole visual scene in detail. Whereas fine-grained vision is only available for the foveated region of the visual field, our brains—lacking supply of detailed visual input from the periphery—may opportunistically extrapolate that peripheral regions must be enriched with the same amount of detail (which, on close visual inspection, generally but unnecessarily proves correct;

see, e.g., Kouider et al., 2012). Thus, in the short run, spatial or temporal continuity in our experience may come about because of a lack of information on discontinuity.

The continuity of conscious experience is also notable on longer time scales. As I am sitting in a Boston hotel room and listening to the squeaks and bell sounds of a streetcar driving by, I am not experiencing a succession of discrete moments, or “snapshots,” but rather a pattern of interwoven, multisensory sensations stretched out and overlapping in time, waxing and waning and interrupted by lapses of attention. One of the prime benefactors of this subjective continuity across seconds to minutes is *working memory*, a short-term form of memory that serves to store stimulus or action information for use in the very near future. It has also been circumscribed as “scratch-pad memory,” as we typically use it to store an item such as a telephone number in memory for a brief time interval—until we reach a phone and dial the number. Working memory depends on the *active maintenance* of item representations, that is, by collective firing patterns of neurons (Goldman-Rakic, 1995; Fuster, 1997). This active maintenance not only takes care of keeping the item in memory after it was presented but also enables other brain mechanisms to manipulate the information and prepare for subsequent action.

Working memory contributes to the sense of temporal continuity by its dual-sided nature: it is both *prospective* and *retrospective* (Fuster, 1997; Rainer et al., 1999). Retrospectiveness is logically implied by any definition of memory because it refers to storage of information acquired in the past. The prospective aspect of working memory is less straightforward and originates from the *association* between past events with stimuli, events, or actions that are predicted to follow next. When, in the early morning, your working memory maintains information that you just loaded your espresso machine, this knowledge instructs you to make the next step toward getting yourself a cappuccino—blowing steam into a can of milk. Sometimes working memory proves its use when a stimulus makes us anticipate a future event that we cannot influence from happening, such as when a flash of lightning elicits the expectation of a thunder—which is in any case less fearful than being fully surprised by an overwhelming, crackling sound. In sum, working memory endows us with a capacity to “chain” past and upcoming events together.

Returning to Harry’s walk in the woods, he did not refer to a role for long-term memory in shaping his conscious experience. Yet, this role is undeniable when it comes to explaining how Harry reacts to the forest scenery. He does not freak out when he finds himself in the forest—he recognizes the trees, and they evoke a sense of familiarity, albeit not at the forefront of his consciousness. If the trees would be *entirely* novel to him, his journey would be a very scary experience: what will these trees do to him? How to get out? Instead, seeing the trees in their colorful splendor has a relaxing effect. He may not have seen any particular tree in the forest before, but the sight of it activates a ramifying set of long-term memories that associate the trees with a vast range of items and properties linked to them. Trees are pretty harmless creatures. They live, but do not bite. They are plants, which are sessile living beings, leaving you alone. Long-term memory, activated by sensory input,

enables us to recognize and interpret sensations in the light of previous experiences. It places the input in a spatiotemporal context that helps us determine our next actions. Much of this associative recall goes on in the background of our consciousness. Even nonverbal, procedural memories may be activated, referring to what Polanyi (1966), Dreyfus (1979), and others consider being part of “tacit knowledge.”

Long-term memory works across much longer time intervals than working memory. We rely on it when recalling our earliest childhood experiences and use it to store information from our current life for the very long future. It also fundamentally differs from working memory by the way it is implemented in the brain. When you are prompted to think of the face of a friend you have not seen for years, visual information should be retrieved from long-term memory and will then be actively maintained in your working-memory “buffer” until further use. Working memory functions as an information buffer with a two-sided Janus face: novel information can enter it transiently, to be passed on to long-term memory. Following time’s arrow in the reverse direction, old information retrieved from long-term memory can enter working memory as well. Roughly speaking, human working memory can hold up to seven plus or minus two items simultaneously in store (Miller, 1956; Lisman & Idiart, 1995). In contrast, long-term memory capacity is vast and virtually without limit. Whereas items in working memory are actively maintained by enhancements of firing rate of neurons in prefrontal, parietal, and connected areas (see figure 2.9), long-term memory mainly rests on the biochemical modifications of synapses (see figure 2.7) in networks distributed across cortical and subcortical structures (McClelland et al., 1995; Fuster, 1997; Frankland & Bontempi, 2005; Battaglia et al., 2011).

3.4 Emotions

What is it about Harry’s autumn forest that made him feel calm and joyful? Intimately connected to memory as emotions are, no agreement exists among psychologists or neuroscientists about what emotions exactly are, but let me sketch the aspects most commonly associated with them. A full-fledged emotion is a complex of sensory, motor, and cognitive processes that all deal with how to respond to, evaluate, and feel about a stimulus or situation that triggers them (cf. Lang, 1994; Frijda, 2006; Grandjean et al., 2008). Different from the long-lasting nature of moods, an emotion is seen as a more immediate brain–body event, caused by a time- and space-discrete source such as the sudden appearance of a spider or a surprise visit by a lover. This “trigger” event is sensory by nature but can quickly evoke a state of motor readiness. A state of motor readiness is expressed by changes in muscle tension but also via the autonomous nervous system, changing our heartbeat, the state of sweat glands, gut contractions, and even pupil diameters, largely outside our voluntary control (Damasio, 2000; Lang & Davis, 2006). In turn, these somatic motor reactions trigger changes in the sensors all over and inside our body, signaling to the sensory systems of the brain that our body state has altered. For instance, when you turn angry,

the cardiovascular buildup of tension is detected by sensors in the aortic arch and more peripherally in the body. These proprioceptive and internal body signals are relayed to the thalamocortical system, where they are centrally represented in the primary and secondary somatosensory cortex (see figure 2.11) but also reach higher-order cortical areas such as the anterior cingulate and insular cortex (Craig, 2002; Pollatos et al., 2005). These brain-body systems underscore the “feeling” (internal sensory) aspect of emotions, which at the same time can influence our further plans and actions as an important variable in itself.

Intertwined with our emotional body reactions and sensations is the cognitive processing of triggering stimuli and the contexts in which they occur. We may be startled for a brief moment by the sight of a snake, but the realization that the snake is behind glass and in a cage shortcuts this early arousal of fear. Our cognitive appraisal of the trigger event may be slower than our initial, more reflexive reactions, but it is no less powerful. By definition, emotions are never neutral. We evaluate the objects in our environment as being relevant or irrelevant. If we find them relevant, we can assign a positive or negative value to them. Positive value can be attributed to an object when it has qualities that we experience directly as rewarding and enjoyable (such as chocolate) or when our brain has learned that the object predicts something pleasant (such as the wrapper of a chocolate bar).

The evaluation of experiential events is inextricably linked to our “feelings” or body sensations aroused by them. Imagine what it would be like to suffer excruciating pain from a knife stab but evaluate the event positively at the same time. In certain pathological conditions, but also theoretically, it is possible to separate evaluation from the experience of body sensations. There is an extensive field of research on robots and artificial neural networks that learn to perform complex motor tasks by so-called reinforcement learning (Widrow et al., 1973; Hertz et al., 1991; Pennartz, 1996; Sutton & Barto, 1998). Here, a neural network processes sensory inputs from the environment and emits motor signals that act on the environment to reach a goal. For instance, the neural network may be part of a robot that learns to chase a “prey” object that, when caught, will allow the robot to acquire energy and recharge its batteries. At the onset of training, the robot will roam around the test arena randomly and cluelessly because its network has not learned yet how to convert the visual inputs of prey movement into motor commands. However, when it accidentally bumps into the prey and gains access to energy resources, a separate, nonvisual input into the network will signal a “positive value.” This has the effect of modifying the connections between the model neurons. The positive feedback changes the neuronal connections such that, on a future occasion, the same configuration of visual inputs will trigger the successful motor behavior more readily. This training technique depends on this reinforcing feedback that instructs the network to adapt—but this by no means implies that the neural network has “feelings” or body sensations. Evaluation and value signaling are key elements in learning and adaptive behavior of both living and artificial organisms, but they should be conceptually dissociated from body sensations or feelings.

I will not try to cover all aspects of Harry's experience or those that others may want to add as elements of consciousness. In Harry's description we find intentions, willful actions, and a selective attention that enables him to "pick out" a tree while neglecting other objects in the scene. He mentions imagery as an important instance of conscious experience distinct from perception. And while listening to Harry, we tend to forget that he is almost constantly using language to describe what he experiences or when pondering about it during his walk. All of these aspects can be taken to contribute to our conscious lives although only some of them will turn out to play an essential role.

3.5 Functionalism

The experiment Harry engages in voluntarily is set against the backdrop of debates on what the most serious problems are when trying to grasp the neural basis of consciousness. An important philosophical direction called *functionalism* posits, briefly, that the mind–brain functions as an entity acting as a *causal intermediate* between sensory inputs and behavioral outputs (Putnam, 1975a; Lewis, 1980a, 1980b; Armstrong, 1980, 1993; Dennett, 1991). Within the larger family of materialistic philosophies, functionalism rejects the conception of mental states as subjective feelings or qualitative experiences in their own right but views them as entities triggered by inputs and exerting particular functions in our behavior. This underlies a kinship between functionalism and behaviorism, which stresses that mind–brain operations should be understood from the study of how stimuli give rise to behavioral outputs, avoiding any introspection and without attempting to dig into the "black box" that lies in between. When Harry's mind is in a state of smelling cinnamon, the olfactory input sniffed up his nostrils, combined with his desire for savory food, will elicit a behavioral tendency in him to pick up the cinnamon can and shake it upside down to cover the butter beans he cooked. The perception of cinnamon is what stands causally in between the olfactory input and the shaking of the can. Therefore it is not a set of perceptual features intrinsic to a subjective state that makes it "mental." The odor perception should be understood by its causal function as it acts as an intermediate between sensory input and motor output. Other mental properties such as beliefs, desires, emotions, and memories should also be explained from their functional relationships with each other and the environment.

This view entails that we should be able to reproduce Harry's sensation of cinnamon if we were to reproduce the functional structure of his olfactory–cognitive–motor system by other elements than axons, dendrites, synapses, and other greasy and slimy stuff. We may as well work with *in silico* or optic components replacing his wet brain circuitry, as long as those parts get the job done of converting olfactory input into can-shaking movements. To make sure that the cinnamon sensation would not be understood as a reflexive or experientially "shallow" or "empty" phenomenon, functionalists argue that the functional structure of the cinnamon-response system be enriched with memories, emotions, inclinations toward decisions, and so on, but in the end, every shade of Harry's joy in cooking and

olfactory sampling should be reproducible in an artificial system implementing the same causal relationships as mediated by his own brain. This latter aspect is called *multiple realizability*: mind–brain functions can be realized in multiple ways, not only by neurons and supportive biological machinery but also by other physical media implementing the same functions as neurons. Traditionally, the functionalist claim that all there is to mental states is their role as causal intermediates in input–output relationships goes hand in hand with its apparent corollary, multiple realizability.¹ However, as I will argue later, it is possible to logically dissociate these two claims, because one can well defend multiple realizability as a principle without having to assume that causal functioning in input–output relationships is all there is to mental states.

Fierce in its rejection of mind–brain dualism, functionalism warns against concepts of the mind as a “Cartesian theater” where sensory data can be watched, felt, or inspected, and where behavioral choices are centrally made. In functionalism, there is no place for a “self” as an active, monolithic agent inside the brain that does all the perceiving and decision making. Dennett (1991, 2003) conceptualized selfhood as a center of narrative gravity. By telling ourselves and our fellow citizens stories about what we see and feel, how our bodies react to events, what we believe and wish to happen, we create a narrative construction to the effect that all of our experiences seem to originate from a single, unified agent we call the “self”—whereas the neural processes underlying all of cognition need not be unified at all.

As functionalists emphasize the importance of studying sensorimotor (input–output) relationships instead of placing subjective reports centrally, it becomes interesting to follow their interpretation of multimodal, perceptual qualities (or qualia). When discussing colors as apparent qualia, Dennett (1991) states the following:

What science has actually shown us is just that the light-reflecting properties of objects cause creatures to go into various discriminative states, scattered about in their brains [...]. These discriminative states of observers’ brains have various “primary” properties (their mechanistic properties due to their connections, the excitation states of their elements, etc.), and in virtue of these primary properties, they have various secondary, merely dispositional properties. In human creatures with language, for instance, these discriminative states often eventually dispose the creatures to express verbal judgments alluding to the “color” of various things. (pp. 372–373)

According to Dennett, the particular “feel” a qualitative feature has to us—the red color of a scarf or a burning toothache—should be interpreted as a discriminative state of neurons in the brain, which has the effect that a sensory input evokes a varied set of emotions, beliefs, judgments, and verbal utterances about that input. The dispositional nature of these secondary properties is taken to reflect our built-in, constitutive inclinations to react to object properties in a characteristic way. The sight of the red scarf will first activate a distributed array of color-sensitive neurons in, for instance, visual areas V4/V4 α (Zeki, 2001) and then trigger pattern changes in higher-order brain areas and engage reactive processes such as emotional expression, intentions, and motor preparation, linguistic

operations, and recall of events and objects we associate with the scarf. If it is true that our conscious experiences are entirely made up of such reactive states that bias individuals toward verbal judgments, it is predicted that experiences should disappear if the subject would be stripped of his emotions, associative memory, motor systems, language capacity, and so on. Let us see what happens to Harry when he asks the neurosurgeon to switch on the neuropharmacological pump.

3.6 What Can Be Peeled Away before Consciousness Is Lost?

Harry is brought into the neurosurgery unit of the hospital. The neurosurgeon applies local anesthetics on his scalp so that tiny holes can be drilled in the skull for intracranial injection of compounds that will inactivate a brain area underneath. The microinjector starts buzzing to inundate Harry's medial temporal lobe region (see figure 2.9) with muscimol—a substance that activates GABA_A receptor channels, inhibiting the firing activity of neurons in the brain, silencing the region for several hours. While structures such as the hippocampus and entorhinal cortex (see figure 2.10) are knocked down one by one, the neurosurgeon interrogates Harry. Across the first hour he politely reports his observations of the surgical ward, the bright operation lamp, the parking lot visible through the window—all business as usual. He reports that his feelings remain as they were—he is a bit tense and worked up, feels a slight itch on his head at the point where the infusion tube enters his skull, but is also curious to see what will happen. But still—nothing happens.

"What do you see and feel now?" the surgeon asks after another half hour.

"I see a nurse wearing a green apron, a bright surgery light shining right in my face, a red car outside," Harry happily continues, and while he jabbers on, the surgeon notices how excited Harry still is, how he goes on tirelessly.

"But Harry," the surgeon interrupts, "haven't you told me this already several times before during the past hour?"

Harry is perplexed.

"What do you mean? I've just entered the experiment and am still waiting for your infusion to get started!"

This kind of effect is what one expects to happen when major portions of the medial temporal lobe are knocked out for longer time periods. We know from studies on patients with permanent brain lesions that this region is crucial for the formation of long-term memories. One of the most famous patients in neurological history, Henry Molaison, who lived from 1926 to 2008 and was known as H.M. throughout numerous publications, underwent a bilateral resection of his medial temporal lobe region in 1953. By this new and intrepid surgery, William Scoville attempted to treat a severe epilepsy Molaison was suffering from. The removal of the hippocampi, plus adjacent regions, from both hemispheres of the

brain resulted in severe memory deficits (Scoville & Milner, 1957). Remarkably, many of Molaison's perceptual, intellectual, motor learning, and language capacities stayed intact. Especially Molaison's capacity to store and remember novel daily-life events (episodes, hence "episodic memory") was destroyed, a condition named "anterograde amnesia." He also had great trouble in recalling events that happened in the recent past before his surgery (retrograde amnesia), whereas he did remember events or facts from many years ago, at least in crude, incomplete, and somewhat abstract form (Corkin, 2002; Moscovitch et al., 2006).

Not all of his learning capacities were affected; he was able to trace figures he could only perceive via a mirror, and he got better at this motor skill across training—but was unable to report in a given session that he had performed the same task on the day before. He showed a dissociation between memories he could consciously recall and talk about—*declarative memory*—and memories of motor skills and other "automated" action patterns—*procedural memory*. Similar cases of patients with temporal lobe damage, often resulting from viral encephalitis, were reported over the decades that elapsed since Molaison's operation (e.g., Stefanacci et al., 2000; Moscovitch et al., 2006; Gold & Squire, 2006). Brain lesion and neuroimaging studies suggest that especially the hippocampal formation is critical for storage and recollection of autobiographical (or "episodic") memories (Moscovitch et al., 2006; Tse et al., 2007; Eichenbaum et al., 2011; Battaglia & Pennartz, 2011). Besides autobiographical memories, which are marked by specific times and places a person visited, declarative memory also comprises the memory for facts and generalized events—labeled "semantic" memory. In many patients suffering medial temporal lobe damage, a relative sparing of semantic memory has been found, although the amount of sparing is still hotly debated (Moscovitch et al., 2006; Squire et al., 2004). But can semantic memory also be impaired while episodic memory is preserved? Indeed a "semantic dementia" has been described, where patients are unable to use and understand general concepts while their autobiographical record is relatively preserved (Hodges & Patterson, 2007).

In Harry's case, a second deficit may have been hidden under his anterograde amnesia. Close to the hippocampus and entorhinal cortex lies the perirhinal region. Arresting Harry's perirhinal cortex by muscimol deprives him, among others, of a sense of *familiarity* when facing known persons or objects (Brown & Aggleton, 2001; Eichenbaum et al., 2007; but see Squire et al., 2004). Being familiar with a particular object one has seen before is different from recollecting the details of that object. Often we see somebody walking on the street or in the subway that we know we have seen before, without recalling her name or other personal details. A memory for familiarity or the sense of "having seen before" is more basic and primary than experiencing the full monty of detailed, associative recall.

How important is long-term memory in shaping our conscious lives? As we walk across the pavement along Main Street, we continuously retrieve details associated with objects we are encountering. Traffic signs are interpreted almost automatically; clothes in shop windows remind us of a current budget limit; a hotdog stand recalls the name and face of

the vendor. If the medial temporal lobe memory system is silenced, conscious life unfolds in a “floating present”—a perceptual moment that is flanked by oblivion about what happened a few minutes ago, and by the abyss of not knowing what will happen next. Mentally invalidating as the destruction of long-term declarative memory is, both amnesic patients and neuroscientists do agree on one major conclusion: *they are conscious*. The richness of their mental life has been diminished by loss of memory, but they still see their environment, perceive sounds, have internal body sensations, and are fully aware of things they touch.

Would it make a difference if, in addition to long-term memory, working memory would be knocked out? Taking Harry’s experiment a step further, his neurosurgeon now infuses muscimol into brain regions deemed essential for this type of memory. Large portions of his lateral prefrontal cortex are now inactivated.² By doing so, the surgeon follows a similar approach as Joaquin Fuster, Goldman-Rakic, and colleagues undertook in the 1980s and 1990s (Shindy et al., 1994; Chafee & Goldman-Rakic, 2000). These researchers reversibly silenced this and other cortical regions by cooling and showed that indeed working memory is hampered. Extrapolating these results to Harry, he now becomes impaired not just in recalling what happened hours to minutes ago in the surgery ward but also in remembering what occurred seconds to minutes ago. His “experiential present” has become narrower in time, and his sense of continuity and time declines (cf. Knight & Grabowecky, 1995; Koch et al., 2002; Harrington et al., 1998). He still responds to the sound of an ambulance’s siren outside the hospital, orients toward it and talks about it, but maintaining memorized information against such a distractor, planning a more complex sequence of movements, or telling an extended, coherent story now becomes very difficult (Ptito et al., 1995; Bechara et al., 1998; Müller & Knight, 2006). Nevertheless, Harry’s perception and his emotional, behavioral, and verbal reactivity to events are largely intact. Indeed, studies on patients with prefrontal or parietal damage have documented defects in working memory but also indicate that a range of cognitive phenomena, including conscious perception, remains intact in the same patients (Pollen, 1995; Knight & Grabowecky, 1995; Bechara et al., 1998).

3.7 Taking Out Emotions

With his memory knocked out, Harry is unable to recall what the next step in his self-imposed experiment will be. However, in accordance with his previous orders, the surgeon switches off Harry’s emotional systems. He activates the pumps driving muscimol into Harry’s amygdala, an almond-shaped complex of nuclei situated close to the hippocampus. This group of nuclei is involved not only in fearful emotions (see chapter 2; Maren & Quirk, 2004; Phelps & LeDoux, 2005) but also in storing memories of positive life events, such as engaging in eating, drinking, and sex, and their environmental predictors (Baxter & Murray, 2002; Parkinson et al., 2001; Paton et al., 2006). The amygdala has often been

credited with a role as the leading actor in our emotional lives, but this would be a simplification. The amygdaloid nuclei each have a distinct set of outputs to other brain structures but are also controlled by other brain areas, especially the prefrontal cortex (Milad & Quirk, 2002; Orsini et al., 2011; figure 3.2). Some regions targeted by amygdala output regulate autonomous body reactions—such as changes in heart rate, perspiration, preparation to put up a fight—whereas others function in higher-order emotional processes such as appraisal and evaluation of stimuli. Particular amygdaloid nuclei are embedded in a network including the orbitofrontal cortex, medial prefrontal cortex, and nucleus accumbens, and collectively these areas are thought to regulate our emotional reactions to rewarding or aversive events and to predictors of these events (think, for instance, of money as a predictor of food; Parkinson et al., 2001; Schoenbaum et al., 2009; Pennartz, Ito, et al., 2011).

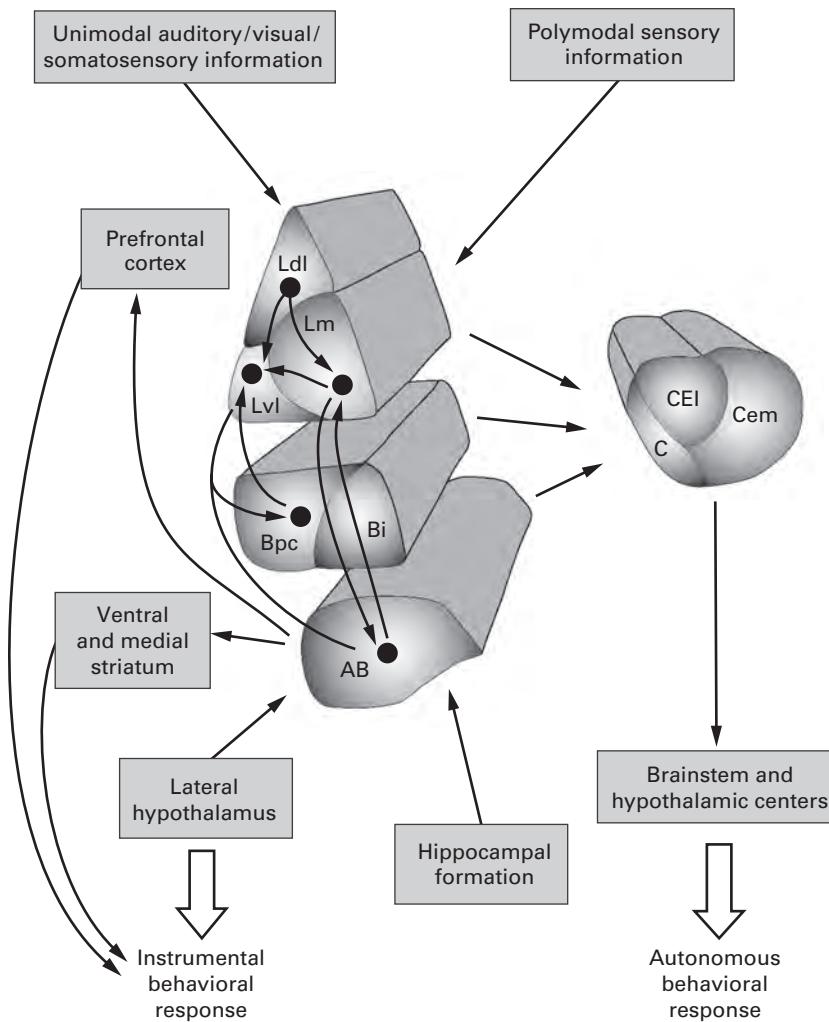
How are Harry's experiences affected when the activity of these emotional brain structures is eliminated? The surgeon has the following discussion with Harry:

"I'm beginning to feel a bit uncomfortable with the situation, Harry. But what do you see now and what do you feel?"

"No problem, surge," Harry answers. "Why uncomfortable? My heartbeat signals jump up and down on the screen here, and that's all fine. I feel that thing on my head where the tubes go in, but the itch does not bother me. I don't really care."

Removing activity in Harry's emotional network makes him indifferent and unresponsive to a situation that would normally elicit strong feelings and reactions. Given that amygdala lesions in monkeys and other primates cause emotional dullness (in addition to other symptoms such as excessive oral behavior and hypersexuality; Horel et al., 1975; Baird et al., 2007), such flattening of Harry's emotionality is plausible although this does not mean that lesions to different individual structures in the network have the same effect. Lesions of the central nucleus of the amygdala in the rat, for instance, hamper very basic expression of learned fear responses, such as freezing and suppression of ongoing behavior in response to a sound that predicts an electric shock (Killcross et al., 1997; figure 3.2). In contrast, lesions of the basolateral nucleus impair choice behavior directed at evading this fearful stimulus. Likewise, damage to our orbitofrontal cortex does not blunt all emotional behavior but makes us inflexible when we have to change our behavior toward items subject to changes in emotional value, such as when we would persist in buying a ready-made meal in the supermarket that made us feel nauseous on a previous occasion.

If neural activity in Harry's amygdala, prefrontal, and cingulate areas and connected striatal regions is arrested, his emotions and body reactivity to arousing events will be blunted, but he will remain conscious of sensory input. There is no impairment in perception per se—except if we would include "emotional feeling" in our definition of perception. Because "feeling" is an ambiguous term, let us analyze what kind of sensations would be lost or not. In our definition of emotion, a stimulus triggers a cascade of brain reactions that result in somatic and autonomous motor effects such as changes in heartbeat and gut

**Figure 3.2**

Anatomical complexity of the amygdala and its main input and output pathways controlling behavior. The amygdala complex is subdivided into distinct nuclei often containing multiple divisions. The nuclei shown here do not represent all subregions recognized but illustrate the main structures implicated in emotional processing of sensory inputs (top) and in expression of behavioral responses and autonomic output. A subgroup of lateral amygdaloid nuclei is implicated mainly in processing unimodal and polymodal sensory inputs. Their outputs are relayed both to centrally located nuclei (right), which affect autonomous behaviors such as freezing and piloerection in rodents, and to the subgroup of basal nuclei. This latter complex communicates extensively with hippocampal, prefrontal, and striatal regions and has been implicated in instrumental, goal-directed behavior such as lever pressing for food. Not all known connections are represented in this diagram, and most connecting arrows reach multiple nuclei within a subgroup. Abbreviations: Ldl: lateral nucleus, dorsolateral division; Lvl: lateral nucleus, ventrolateral division; Lm: lateral nucleus, medial division; Bpc: basal nucleus, parvicellular division; Bi: basal nucleus, intermediate division; AB: accessory basal nucleus; C: central nucleus, capsular division; CEI: central nucleus, lateral division; CEm: central nucleus, medial division. Based on Savander et al. (1997), Groenewegen and Uylings (2000), and Cardinal et al. (2002).

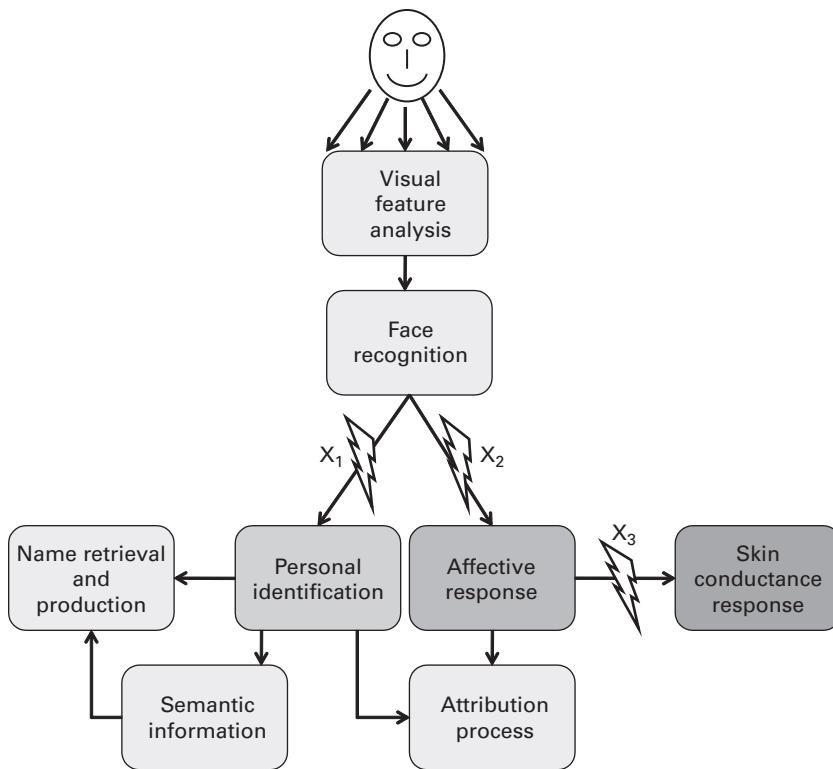
contraction. These body reactions are signaled by sensory transducer cells in our muscle, skin, and vascular tissue and relayed to the sensory systems of the brain. In that sense, a silencing of the emotional network *will* lead to a change in perception, but this only pertains to the body sensations evoked by brain–body reactions to the initial stimulus—not a change in the exteroceptive senses such as vision and audition, which dominate conscious experience in humans. Harry does report on a tactile sensation around the drill hole in his skull, but he is indifferent to it. Normally he would have felt anxious about this self-imposed infliction, resulting in intense perspiration and other body reactions, but because the central brain nodes regulating these reactions are no longer functioning, these emotion-related sensations are lacking. Apart from this, Harry is fully aware.

A poignant illustration of the dissociability between emotion and perception is the syndrome of Capgras. Patients suffering from this rare, delusional condition can recognize family members or other familiar persons but believe that they have been substituted by impostors. They lack an emotional, affective response to seeing and hearing these significant others (Ellis & Lewis, 2001; Devinsky, 2009). Ellis and Lewis (2001) describe how a male patient, having suffered from a car incident, believed that his wife had died in the crash and that the person he was living with was a deceptively realistic replacement. Patients typically lack the *feeling* that their nearest ones are the persons they know and love—the connection between perception, on the one hand, and, on the other hand, emotional recognition and the spark of familiarity is missing. The neuroanatomical basis of this syndrome is largely unknown, but it has been hypothesized that the patients have intact temporal lobe memory systems as well as functioning emotional structures but suffer from a deficient communication between these systems (cf. Breen et al., 2000; Ellis & Lewis, 2001; figure 3.3).

Now Harry's experiment has arrived at a point where we can begin to discern theoretical implications. Functionalism posits that the quality of an experience should be understood as a specific discriminative brain state acting as a causal intermediate between sensory input and behavioral output. The particular “feel” or subjective experiencing of a color would come about because the structures of our color-response systems include, or interact with, connected brain systems for personal memory, emotions, language, and motor responses. However, the evidence from neurology and neurobiology indicates that qualitative experiences remain largely intact when neural coding of personal memories and emotions is eliminated. The experiences may be impoverished by their lack of emotional color and historical depth—but the vision of color and the smell of cinnamon are still there. Let us next see what happens when our brain systems for motor behavior and language would be blocked.

3.8 Consciousness in the Absence of Motor Behavior and Language?

One of the perplexing aspects of mind–brain research is the abundance of opinions and definitions of consciousness, with even more background assumptions. Whereas this book

**Figure 3.3**

Model of processes disrupted in Capgras syndrome (based on Breen et al., 2000, and Ellis & Lewis, 2001). Visual input is analyzed via feature detectors in the visual system, resulting in face recognition and (via the left branch) identification of the person the face belongs to and retrieval of his name. A defect at position X_1 but not X_2 and X_3 results in a disability in putting names to faces, corresponding to prosopagnosia. A defect exclusively at X_2 may explain Capgras syndrome as the affective response to the face is lost but the person can still be identified, which leads to a mismatch in a higher-order process attributing the face to a “look-alike.” A lesion at X_3 leads to deficits in autonomic emotional responses to stimuli such as measured in skin conductance. Such deficits can occur in patients with lesions of the ventromedial frontal cortex. Also a deficit at X_2 can result in a loss of skin conductance response, consistent with Capgras syndrome.

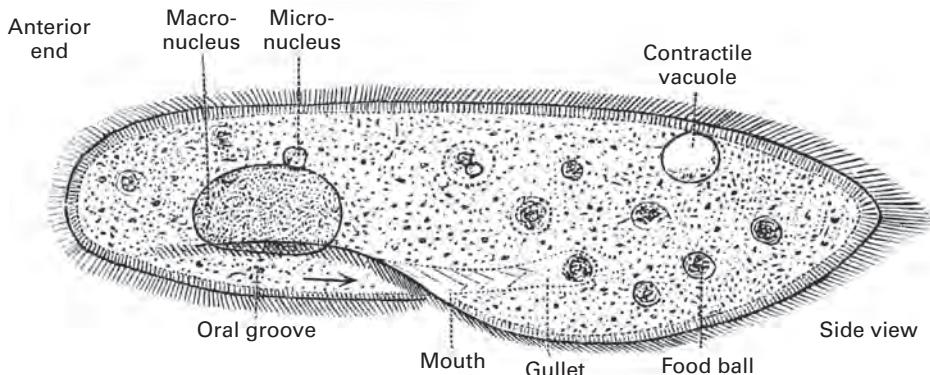
argues that sensory processing is key to understanding consciousness, others view it primarily from the angle of free will and motor behavior. Llinás (2001), for instance, emphasizes the importance of active movement in cognitive processes such as self-representation: “Ultimately we see that the architecture capable of generating cognition must relate to the motricity upon which such cognition was developed” (p. 265).

This position is related to arguments on the evolution of consciousness. Did consciousness arise gradually along the course of evolution, as simple organisms evolved into more complex species—or did a sudden “jump” from nonconscious to conscious systems occur? Scientists in favor of such a nonlinear, Big Bang-like step often point to the period when

hominids started using language, perhaps some 100,000 to 50,000 years ago (Mithen, 1999; cf. Carruthers, 1989; Jaynes, 1990; Arbib, 2001).³ Llinás develops his account of motricity-driven cognition by considering primitive creatures called “sea squirts” (*Ascidacea*), which spend their adult life in a sessile state, with their pedicles safely fixed to a rock on the seabed. Whereas the adult animal harbors no nervous system except for cells regulating its filter-feeding behavior, the preceding stage is a larva that has a tail and swims around freely—and has a brain-like ganglion and attached spinal cord-like structure (Barnes, 1980). This primordial nervous system enables the animal to process information from skin photoreceptors and a balance-sensitive statocyst and to transduce this sensory input into movement. Llinás’s take-home message, supported by other examples, is clear: active movement is an evolutionary prerequisite for having a brain and mind. But how important exactly is active movement for conscious experience? We follow the same evolutionary lead here and ask whether the combination of sensory signaling and motor output would be sufficient to label an organism conscious. Take a friendly organism called *Paramecium*, a single-celled creature with a slipper-like shape, about 0.25 mm long and found swimming around in ditches (figure 3.4). Covered as this creature is with *cilia*—hair-like flagella producing motion by undulating movements—it can recoil when touching a solid surface, putting its motion engine in reverse, change direction, and head forward again. Llinás (2001) notes:

We know that single-cell “animals” are capable of irritability, that is, they respond to external stimuli with organized, goal-directed behavior. [...] And so we are left with the nagging feeling that irritability and subjectivity, in a very primitive sense, are properties originally belonging to single cells. If so, that primitive subjectivity may be built into the consciousness and subjectivity displayed by the nervous system as cells organize into the ensembles that we know as nerve cell circuits. (p. 113)

Of the many questions this passage raises, only a few points will be discussed. It can be debated whether *Paramacial* behavior is organized and goal-directed in the sense that it would transcend a limited set of stimulus-response reflexes, which are fixed and stereotyped. Capacities for planning, goal representation, and organizing complex behavior are lacking. However, such arguments are besides the main point here—the key question is whether a primitive kind of subjectivity may be ascribed to this organism. Clearly, *Paramecium* is irritable—its outer membrane harbors mechano- and chemosensors, and it reacts to environmental inputs—but is there also positive evidence for a subjective, conscious experience in this organism, no matter how primitive? I could not find convincing evidence in Llinás’s account or elsewhere, and the thesis of this book states that primitive organisms like *Paramecium* lack the representational power to generate anything that could be classified as a conscious experience. No evidence is offered to suggest *Paramecium* would be capable of constructing a perceptible scene, binding sensory features into integrated objects, or even identifying what the input, signaled at its outer surface, is about. Maybe

**Figure 3.4**

Side view of the one-celled organism *Paramecium* (from Colton, 1903). This organism can move forward, move backward, or turn by the use of its cilia covering the cell membrane. The length of this organism is in the order of 250 µm.

these complex operations are not needed to achieve the kind of primitive subjectivity that *Paramecium* has? The perspective here is that any potential “subjectivity” vanishes into the ultimate primitive organization of the organism.

Thus, positive evidence is lacking to accept *Paramecium* as being conscious, as could be obtained through behavioral observation or studying the capacities of its nervous system. With the current state of technology, it should be feasible to produce an artificial *Paramecium*-like organism with the same kind of reflexive sensorimotor system as living exemplars have, but without any need to invoke “consciousness” or cognition. As a cornerstone of science, Occam’s razor will do its work: only include a term in your vocabulary when the explanation demands it. But let us now jump from *Paramecium* back to Harry and examine whether active movement, motor planning, and verbal behavior are crucial for his conscious experience.

We left Harry in a state largely deprived of emotions and recent memory. Ignorant about the foregoing events, yet relaxed and unmoved, he is still seeing, hearing, and feeling and is able to talk about these sensations:

“I am now going to bring down the motor centers in your brain, Harry. Normally this could make people a bit scared, but I will leave the inactivation of your emotional centers as it is, so I don’t expect you to panic.”

“But wait, doc! I vaguely recall something in our protocol about getting my memory system back online. What was it again?”

The surgeon frowns and realizes he forgot an important detail of the procedure:

“You are awesome, Harry. Your temporal lobe memory system is practically shut down, but you still recall this. It must be because you went over the protocol so many times.”

"Well, thanks, doc. But what was the point again?"

"Let me show you."

The surgeon presses a couple of buttons so that the pumps stop infusing muscimol into Harry's temporal lobes, prefrontal area, and parietal area, and he starts an episode of *The Simpsons* in case Harry gets bored. It takes several hours for the drug to be removed, but slowly Harry's amnesia is reversed. Meanwhile another drug pump is activated, infusing muscimol into a distributed collection of motor areas via a ramifying set of tubes. It first takes out the large expanse of primary motor cortex (Brodmann's area 4, just in front of the central sulcus; see figure 2.9). Next down are Broca's area, strongly involved in articulation of speech, and Wernicke's area, involved in analyzing acoustic inputs and comprehending sentences, as well as further language-related areas. The premotor cortex, supplementary motor areas, and frontal eye fields are inactivated, removing capacities for motor planning and selection, complex movement sequences, and voluntary eye movement. Finally, brain regions for learning and executing motor skills and habits, such as the dorsolateral striatum and several thalamic nuclei, are inactivated. And there Harry lies, motionless and speechless. He is now completely paralyzed, cannot even move his eyes to express what he is looking at. He cannot even articulate to himself, let alone to other people, what his state is like. But is he still conscious?

An influential branch of mind–brain philosophy argues that language, social interactions, and consciousness are strongly, if not inextricably, intertwined (Carruthers, 1989; Jaynes, 1990; Jackendoff, 1999, 2007; cf. Dennett, 1991; Arbib, 2001). A first problem in this account is that, on an evolutionary time scale of millions of years, consciousness would have arisen quite suddenly, when mankind developed grammatically structured, verbal behavior. Were humans or humanoids unconscious before this development? Did they suddenly change from "zombies" to conscious beings when they already had sophisticated perceptual, motor, emotional, and memory capacities, but in addition started to speak? Where, along the scale of structured vocalizations, would the border lie between simple vocalizations and consciousness-generating language? Although these questions have prompted proponents of a language-consciousness account to express a wealth of viewpoints, they already raise doubts about a language-based account of consciousness, and more is to come.

Global aphasia occurs when a massive stroke hits multiple language-related regions, including Broca's and Wernicke's areas as well as the basal ganglia, insula, and superior temporal gyrus. Aphasic patients cannot articulate or comprehend structured sentences but are able to carry out linguistic routines such as counting and singing fragments of songs. Suffering from damage mainly in the left hemisphere, they are also paralyzed on the right side of the body, but otherwise their body language is intact, and they show emotional prosody in their remaining speech as well as facial expressions (Blonder et al., 2005). In neurology, there is no doubt that aphasic patients are conscious, at least if other

cognitive and sensorimotor faculties remain functional (Weir et al., 2003). Their repertoire for expressing behavioral signs of consciousness, via eye, head, limb, and facial movements, is sufficiently broad and rich to raise ample evidence they are conscious. Therefore it is reasonable to assume that linguistic–propositional brain systems are not necessary to sustain conscious representations.

Gazzaniga and colleagues (1987; Funnell et al., 2000) presented evidence for the reverse case: linguistic expression persisting in the absence of consciousness. This result reinforces the dissociability between the two phenomena even further. They described a patient whose left and right hemispheres were separated by transection of the corpus callosum (a commissurotomy). If a stimulus is presented in the left half of this so-called split-brain patient's visual field, information will be transmitted to the visual system of the right hemisphere, because of the decussation (left–right crossover) of the optic nerves and the predominant projection of the leftward-oriented parts of the retina to the right hemisphere. Despite the commissurotomy and given specific test conditions, information was still transmitted from the left visual field to the left-hemisphere center for linguistic interpretation and articulation, leading the patient to speak about the stimulus presented. Strikingly, however, the left hemisphere showed no evidence of being consciously aware of processing the item information received from the right half brain.

Going back to the dissociability of motricity and consciousness, it has long been known that a pharmacologically induced paralysis is compatible with a retention of consciousness. Take, for instance, the horrifying experiences of people hit by curare-dipped arrows in South America. Chronically paralyzed patients often retain some ability to report their experiences and thoughts, often verbally. In locked-in syndrome, the means of expression are severely reduced: the patient can voluntarily move one or both eyes, but nothing else. There is little debate on whether locked-in patients are conscious—they are, as verified roughly by the same means we use to identify states in other, healthy people. The interesting point to make about paralysis, however, is that the severity of motor loss is not correlated in any way to the extent to which consciousness is preserved. There is no evidence to suggest that a patient's consciousness would gradually decrease as motor loss becomes more severe. Jean-Dominique Bauby's autobiographical book *Le Scaphandre et le Papillon (The Diving Bell and the Butterfly)*, 1998) powerfully testifies that a person can remain fully conscious while motricity is almost completely eliminated, underscoring that the two are not directly or immediately related. In chapter 11 we will look at the even more extreme case of patients who show brain activity indicative of consciousness in the complete absence of motor activity—namely, in a vegetative state (Owen et al., 2006; Cruse et al., 2012).

Neurology also presents examples of the reverse condition: continued motor behavior in the absence of consciousness. During a specific type of epileptic seizure often reported in children, *absence*, patients suddenly cease behaving, freezing their movement pattern, maintaining posture, their faces blank and devoid of emotional expression. During

an ensuing episode of *absence* automatism, usually lasting seconds to tens of minutes, patients start moving again, performing complex actions such as drinking, walking, putting a shirt on and taking it off, and switching the lights on and off, thereby reacting to environmental stimuli—but still with an empty, vacuous expression on their faces (Panayiotopoulos et al., 1989; Damasio, 2000; Yoshinaga et al., 2004; Sadleir et al., 2009). Following the *absence*, patients are highly surprised upon finding themselves in the situation they brought themselves into—apparently conscious of their situation only after the episode ended, but without any recollection of what went on during the seizure. The lack of conscious recollection may be caused by a memory deficit or by an utter absence of consciousness (or both), but a couple of observations suggest the phenomenon cannot be explained by a memory deficit alone. First, during the *absence* patients lack behavioral reactivity to salient stimuli and rigidly maintain a void expression on their face. Second, a heavily compromised memory system would not be expected to recover instantaneously after an epileptic episode, but patients are fully able to store and recall what happened just after they recovered. The overall syndrome is consistent with motor behavior being intact but conscious perception being absent. This absence of consciousness cannot be claimed in a strong way because we must rely on the absence of behavioral indicators of consciousness (including verbal reporting) during epileptic episodes. But next to behavioral indicators, neurophysiological recordings have shown that *absence* seizures are expressed by abnormal, low-frequency “spike-and-wave bursts” in scalp EEG, which override the normal electrophysiological state of conscious processing (Yoshinaga et al., 2004; see chapter 7).

Looking at Harry’s condition that is not to be envied, we are still facing the question of his consciousness. Unable as he is to squeak, pull faces, or make gestures, his body is completely mute, except for his heartbeat and respiration. If conscious at all, he is completely locked in. Does he sense any hint of reproach or even anger? Harry fails to give a blink. Struggling as the surgeon is with feelings of guilt and remorse about the dubious situation he got himself into, he turns off the pumps and lets Harry recover. An agonizing hour passes. Then, some grimaces gain in strength:

“Hmmm … the strangest thing was happening, doc. I was alive, but felt like dead.”

“What do you mean? Were you angry as the paralysis set in? I *did* try to shut down your emotional system!”

“That wasn’t the problem. I was not angry, not happy either, not in pain. I just watched Homer drive around Springfield. But the experience was completely afloat—I could not react to anything, could not laugh, and the whole thing did not matter to me.”

“But you do remember it all along.”

“Sure enough, *I* was there all along. I saw the movie, was aware of the bed I am lying on, and felt my heartbeat go up and down. But it all felt less like being ‘me’ as so many of my abilities were taken out.”

Where does this leave us, peeling off our conscious lives? Harry’s fictitious case can be further underpinned with further, often bizarre neurological cases. Goodale et al. (1991)

described a patient who failed to perceive object properties such as shape, size, and orientation. Nonetheless, she was able to accurately guide her hand and fingers to grasp these objects—in a way that requires knowledge of these very properties. Brugger et al. (2000) reported a striking case of a woman’s having conscious sensations in her arms and legs without ever having had a chance to move these limbs or learn how to use them—because she was in fact born without them. In conclusion, we do need our motor apparatus to express to others what we perceive, but motor (or linguistic) activity is not necessary for consciousness *per se*—and as long as we retain memory capacities, we can report our experiences afterward.

3.9 Which Brain Structures Cannot Be Peeled Off?

Somewhat incredulously, Harry is staring at the white plastered wall of the surgery ward:

“I can’t believe we went through all this trouble, doc, and we still have not been able to knock down my consciousness,” Harry sulks.

“Well, what did you expect? I suggest we keep your brain systems for memory, emotions, and motor behavior online now, while we take out some sensory areas one by one.”

Harry demurs some more but, in the end, succumbs to his own curiosity:

“Okay, but this time I really do expect some more fireworks, doc.”

“That’s likely to happen, except that you might not know what you are missing.”

The surgeon directs muscimol into Harry’s middle temporal area—in short, area MT or V5 (see figure 3.5 for its location in macaque brain). Harry picks another movie to watch, *The Dark Knight*. He likes action movies but soon realizes that much of the “action” is missing, whereas he does recognize Batman and the Joker. They change positions but seem frozen on the screen. The sound of mobs and SWAT teams approaching conveys a sense of auditory motion, but visual movement has largely disappeared.

The inactivation of MT/V5 (and possibly surrounding areas in the lateral occipital cortex) creates a rare neurological condition called akinetopsia, first described in a relatively selective form by Zihl and colleagues (1983) in a female patient who also suffered from mild aphasia and difficulty with mental arithmetic because of a vascular disorder. Such patients see fountains as if frozen and have trouble following conversations because they cannot perceive the lip movements of the person they are talking to. The syndrome strikingly illustrates the functional dissociations that can occur when some areas of the visual system are damaged while others are spared: inactivating Harry’s MT/V5 area degrades *only* his motion vision whereas his perception of luminance, color, contour, and form remains intact. His conscious vision is intact except for the loss of one submodality or attribute (Zeki, 1993, 2001). Area MT/V5 is part of the extrastriate visual cortex, and if it is inactivated, the lower-level visual areas such as V1 and V2 keep on processing inputs

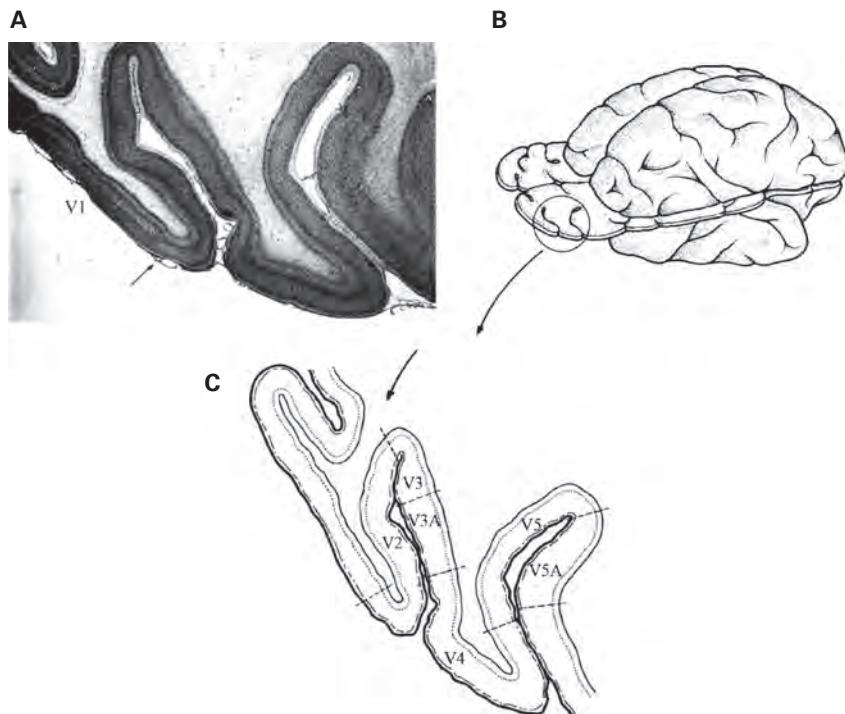


Figure 3.5

Location of the primary visual cortex (V1) and prestriate cortex in the macaque brain. (A) Based on the histological structure of the cortical layers, area V1 has been demarcated from the more anteriorly located prestriate cortex (an arrow marks the boundary). The photograph represents a horizontal section through the occipital lobe, illustrated in (B). (C) Location of areas V2, V3, V4, and V5 (and their subdivisions) along the gyri and sulci also shown in (A) and (B). V5 is alternatively named “MT” (middle temporal) and is strongly involved in motion vision, whereas area V4 is concerned with color vision. Depending on nomenclature, the term “prestriate cortex” can refer to areas V2, V3, V4, and V5 (Zeki, 2005), but often V3, V4, and V5/MT are denominated as extrastriate cortex. From Zeki (2005), reproduced with permission (copyright Royal Society).

from the LGN. Likewise, other higher visual areas, such as V4 and IT cortex, remain intact, and therefore it is possible that only a very selective “loss of consciousness” occurs, limited to one submodality.

The converse case, where patients are blind in part of their field of vision (a scotoma) but still see motion, has also been documented, beginning with Riddoch’s studies on the effects of gunshot wounds in British soldiers during the First World War (Riddoch, 1917, cited in Zeki, 1993). A scotoma is usually caused by damage to a subregion of area V1, and the larger the size of the lesion, the greater the portion of the visual field the patient becomes blind to. Riddoch’s war victims were blind to objects located in their scotomatous fields, unable to report their shapes, color, or contrast to background—except when these objects

started to move. This movement only resulted in perception of motion, not in the additional emergence of form or color of the moving object.

This pattern of selective loss and dissociability is also seen in other visual submodalities. Achromatopsia is the inability to see our surroundings in color. Generally, patients are able to see the shape, motion, brightness, and surface texture of objects as far as these features can be discerned independent of their color. They live in a world painted in gray tones as vividly illustrated by an artist, Mr. I, who became achromatopsic after a car accident (Sacks, 1996):

“You might think,” Mr. I. said, “loss of color vision, what’s the big deal? Some of my friends said this, my wife sometimes thought this, but to me, at least, it was awful, disgusting.” [...] He knew the colors of everything, with an extraordinary exactness [...]. He could identify the green of van Gogh’s billiard table in this way unhesitatingly. He knew all the colors in his favorite paintings, but could no longer see them, either when he looked or in his mind’s eye. (pp. 6–7)

Our neuroanatomical knowledge of conscious color vision is far from complete, but achromatopsia is known to arise usually from damage to the cerebral cortex, unlike conventional color blindness, which originates from defects in cones, a color-sensitive type of photoreceptor in the retina. Crucially, in achromatopsia wavelength information is processed in the retina and subcortical stages, whereas color perception is lacking (Cowey & Heywood, 1997). Although not undisputed, the weight of evidence suggests that achromatopsia is caused by lesions of area V4/V4 α , which is part of the fusiform gyrus in the human brain (see figure 3.6, plate 1). Deficits may also arise when damage occurs to specialized color-processing channels of the lower visual areas, V1 and V2 (the “blobs” of V1 and thin stripes of V2), which relay color information to V4 and higher areas, such as IT cortex (Zeki, 1993; Heywood et al., 1996; Conway & Tsao, 2006; Banno et al., 2011). Frequently, V4/V4 α lesions are of vascular origin and are accompanied by functional loss in neighboring regions in the fusiform gyrus, which is important for vision and recognition of shape and faces (Gauthier et al., 2000). Nonetheless, a fair number of patients lose color vision but retain form vision, underscoring the point that color vision is mediated by a separate subsystem or “node” in the visual system.

Can color vision be spared in the absence of visual perception of other attributes? Victims of carbon monoxide poisoning, mostly due to heavy smoke inhalation during a fire, provide an affirmative answer. Some of these patients have profound visual impairments as they cannot see and report shapes of objects—except that they can name their colors and can group together tones or shades of the same color (Zeki, 1993). Humphrey et al. (1995) describe one such patient who was unable to tell visual grating stimuli apart with a 90° difference in orientation but could distinguish different colors reliably.

Where do these neurological cases leave us in the debate on functionalism? Let us consider the functionalist claim again that a qualitative, conscious experience is constituted by a discriminative, reactive brain state that will bias the subject toward a verbal judgment or

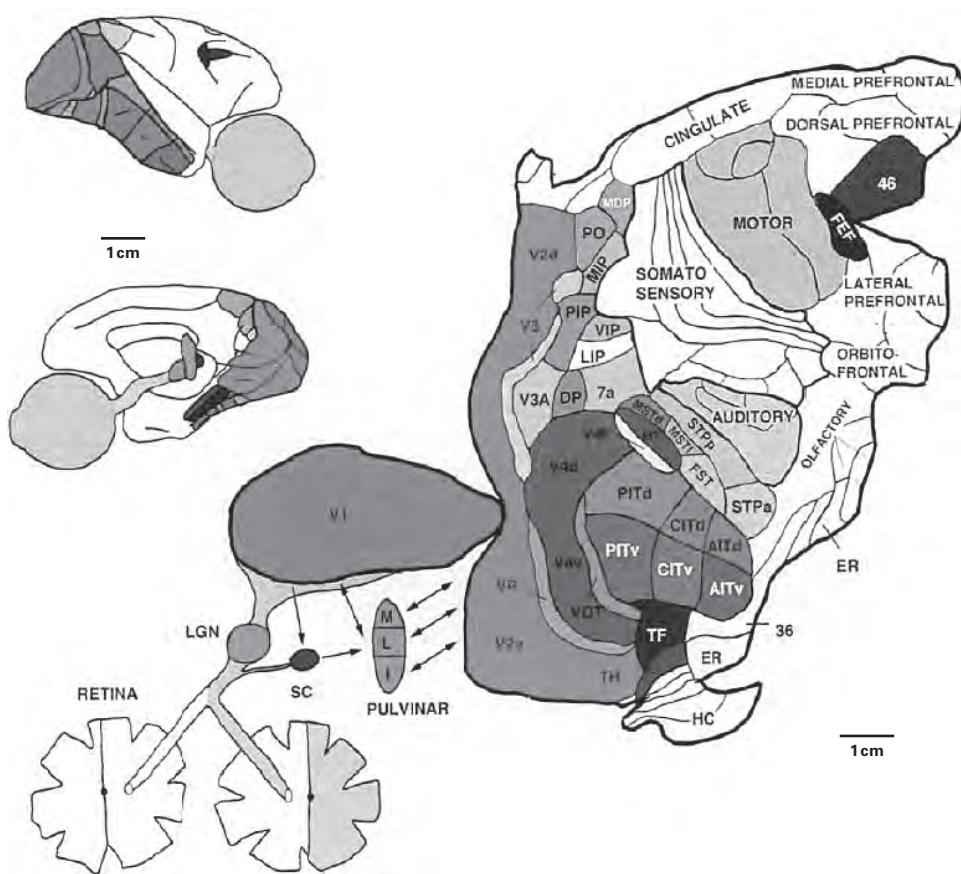
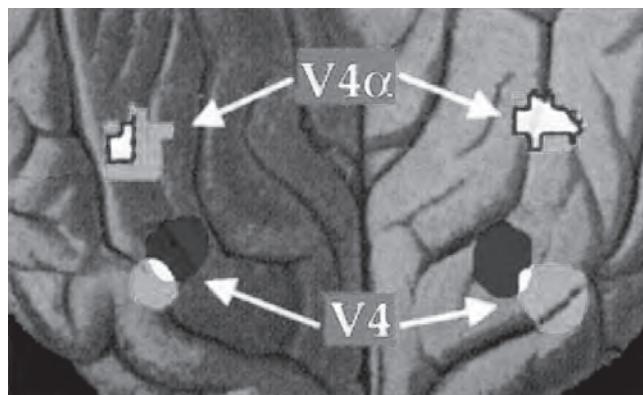
A**B**

Figure 3.6 (plate 1)

Overview of the macaque visual system and areas activated by color vision in the human brain. (A) Top and middle left insets show lateral and medial view of the right hemisphere of the macaque brain, respectively. The middle and bottom panels show the eye (gray), optic tract (light blue), lateral geniculate thalamic nucleus (LGN), superior colliculus (SC), and pulvinar as subcortical visual areas. From bottom to upper right: unfolded map of cortical visual areas shown in conjunction with subcortical structures. The three-dimensional layout captured in the left insets has been projected here on a flat surface, resulting in an artificial discontinuity between V1 and V2. From Van Essen et al. (1992), reproduced with permission from AAAS. (B) Ventral view of the V4 complex in the human brain with superimposed activation patterns derived from brain imaging. Area 4 shows evidence for a retinotopic organization, with the lower part of the visual field represented in green and the upper part in red (yellow indicates overlap between red and green). Also the more anteriorly located area V4α is activated by colored patches, set within a multicolored scene, but does not show a clear retinotopic organization. From Zeki (2005), reproduced with permission (copyright Royal Society).

some other body-language sign. Defying functionalism, would it be possible to obtain a correct verbal response from a subject about a stimulus difference imperceptible to him? Heywood and colleagues (1991) described an achromatopsia patient who was able to pick out an odd stimulus which differed from others only in color and not in brightness (i.e., they were isoluminant). Perceptually, these stimuli were indistinguishable to this patient. The patient declared he was able to detect boundaries between two stimuli but without perceiving a difference between them. Whether the cause of this judgment ability resides in a sparing of V1 or other visual regions remains debated, but apparently it is not sufficient for conscious color perception to have, in our brain, simply a “discriminative state” that allows us to make a correct behavioral judgment. For perception, more appears to be needed than having a bunch of neurons in a state where they *discriminate* in their firing rates between a stimulus A versus B—and relaying this differential response to motor centers. This important conclusion is also confirmed by achromatopsia in general because in this condition loss of conscious color vision coexists with retention of color-discriminating cells in the retina and thalamus and can occur even when areas V1 and V2 process color information (Heywood et al., 1996).

The case of achromatopsia with intact judgment is somewhat analogous to the well-known phenomenon of blindsight, where patients are blind at least in part of their visual field because of damage to their primary visual cortex. However, when prompted, they can guess above chance level where in the visual field an object is or whereto it is moving—despite the fact that they are not conscious of the object presented (Weiskrantz et al., 1995).

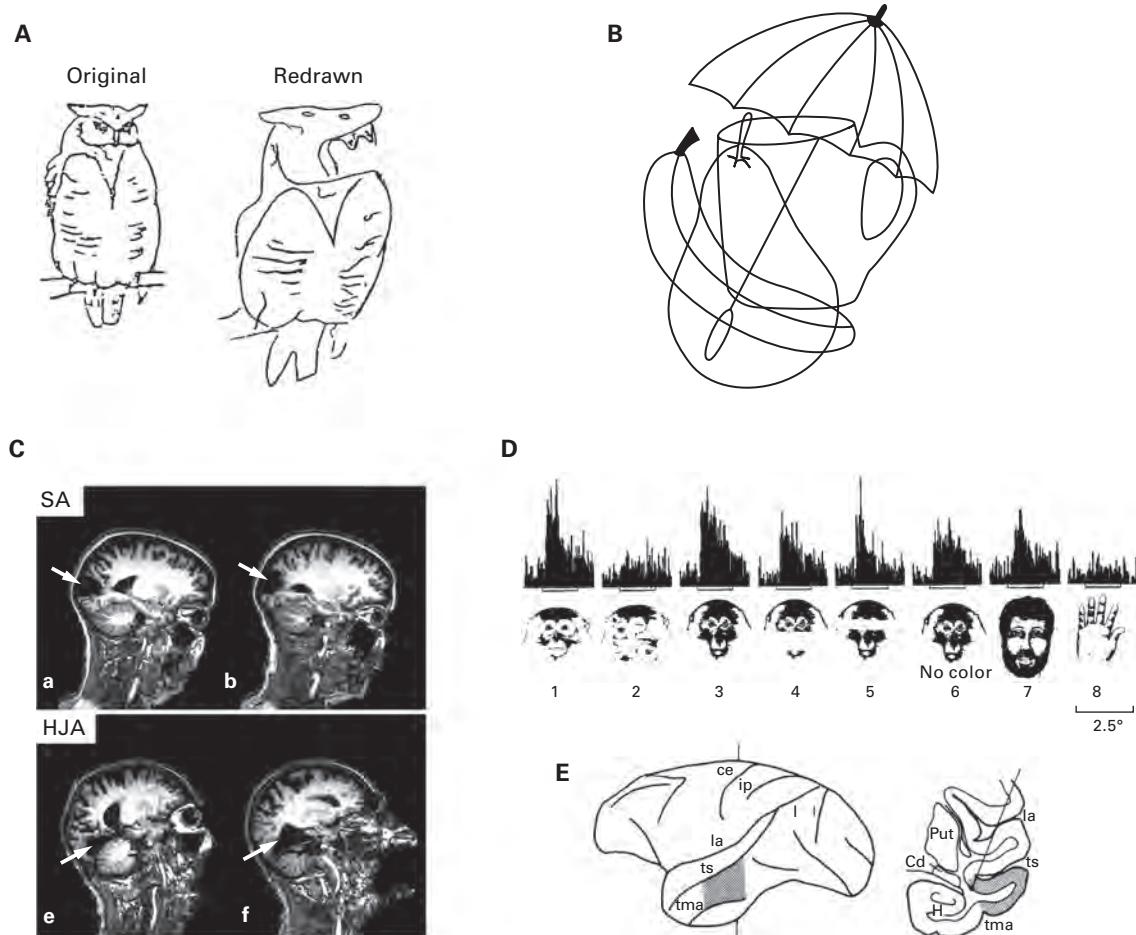
Space is lacking to treat all of the osias and opsias at length, but let me briefly summarize some further deficits that are particularly important for understanding the interface between cognition and perception. *Visual agnosia* is a broad term for syndromes found in patients who are able to process visual stimuli but are unable to recognize them (Farah, 2004). In this case, recognition does not refer to knowledge of the familiarity or novelty of an observed object but to the ability to identify the presented object by naming and describing it. Classically, the agnosias are split up into two classes: “apperceptive” and “associative” deficits (Lissauer, 1890). In apperceptive agnosia, the recognition deficit is ascribed

to a failure in visual perception, whereas in associative agnosia, patients can see objects normally but cannot retrieve names or other properties associated with them (“perception stripped of meaning”; Teuber, 1968).

Visual agnosia can be a syndrome of fundamentally not “knowing” (corresponding to the Latin verb “*gnoscere*”), “understanding,” or recognizing objects, or conversely of not being able to “see” a complex object, as being more than a collection of loosely identified elements.

Nowadays, neuroscientists emphasize the large heterogeneity of the agnosias and do not maintain a strict separation of them in two classes (Zeki, 1993; Farah, 2004). For instance, a patient may be able to copy a complex object such as a building with pen and paper without being able to name or describe the object—a finding which may be interpreted as an ability to perceive, but not grasp the meaning or identity of objects. However, the distinction between “seeing” and “understanding” starts fading away once it becomes clear that many patients copy the picture by bits and pieces. A patient described by Humphreys and Riddoch (1987) declared that, while drawing complex figures such as a building, “all he saw was a complex pattern of lines, which did not correspond to a particular object.” In this type of visual agnosia, which might be traditionally classified as “associative” but is better labeled as an “integrative” agnosia, subjects can identify and use local, low-level features of an object or scene but are incapable of integrating these into a larger perceptual entity (Riddoch et al., 2008; figure 3.7). Magnetic resonance imaging (MRI) scans in visual agnosia patients have revealed lesions in dorsal extrastriate areas and ventromedial occipitotemporal regions, such as the inferior temporal gyrus and fusiform gyrus (Farah, 2004; Riddoch et al., 2008). In monkeys, electrophysiological single-cell recordings have been made from a functionally and anatomically similar area, the IT cortex (see figure 3.7). IT neurons generate spike responses that can be highly specific for complex visual shapes such as objects, animals, and faces (Desimone et al., 1984; Perrett et al., 1992; Gross & Sergent, 1992). For the fusiform gyrus and homologous areas, there is little evidence to maintain a rigorous distinction between “seeing” and “knowing,” leading us to think of the visual agnosias more as a graded mixture of “agnosia” than purely as agnosia by its Latin origin.

As a special form of visual agnosia, prosopagnosia is a more specific deficit: an inability to consciously recognize faces (i.e., identify persons the face belongs to), even when these were familiar before brain damage occurred. The specificity is apparent from an intact visual recognition of nonfacial stimuli and normal level of intellectual functioning (Ellis & Lewis, 2001; Farah, 2004). Patients do know what a face is, as a concept or perceptual category, and how it is composed of ears, eyes, nose, and mouth. However, they cannot visually identify a face as belonging to somebody familiar, such as a friend or a relative. Pallis (1955) cites a patient who tried to identify medical personnel: “You must be a doctor because of your white coat, but I don’t know which one you are. I’ll know if you speak” (p. 220). When his wife came to visit, he failed to recognize her. Despite such

**Figure 3.7**

Visual agnosia in relation to functional properties of neurons in the inferior temporal cortex of macaques. (A) A patient with visual agnosia was able to produce the drawing shown on the right when inspecting the original on the left. Nevertheless, the patient could not describe or name the object that was being copied. (B) One of the neuropsychological tests to study visual agnosia is Ghent's overlapping figure test. The task is to identify all objects included in the composite figure. (C) Magnetic resonance images of brains of two visual agnosia patients (S.A. and H.J.A.) showing large lesions in dorsal (S.A.) and ventral extrastriate cortex (H.J.A.), respectively. Arrows have been added to the original to highlight massive lesions (from Riddoch et al., 2008). (D) Neurons in the inferior temporal cortex of the macaque brain can show highly specific action potential responses to faces. Responses to monkey or human faces can be compared to responses to a hand or to partially occluded or scrambled faces. The bar under each histogram indicates a stimulus presentation period of 2.5 seconds. Each stimulus (1–8) was repeated several times. The height of each vertical bar in the histograms, all derived from recordings from the same neurons, represents the rate at which spikes were fired in response to the visual stimulus. (E) Recording sites in (D) are indicated by the stippled area in a lateral view (left) and a coronal section through inferior temporal cortex (right). Neurons were recorded using metal (tungsten) electrodes. One of the electrode trajectories is illustrated by the line from top right into the stippled area on the right. ce, central sulcus; ip, intraparietal sulcus; la, lateral sulcus; ts, temporal sulcus; tma, anterior middle temporal sulcus; H, hippocampus; Cd, caudate nucleus; Put, putamen. Adapted from Humphreys and Riddoch (1987; A), Giannakopoulos et al. (1999; B), Riddoch et al. (2008; C; reproduced with permission from Taylor-Francis); (D) and (E) from Desimone et al. (1984), reproduced with permission from the *Journal of Neuroscience*.

identification deficits, some patients are still able to discriminate the emotional expression on a face, be it full of joy or sadness. They may even reconstruct familiar faces by imagery, but they cannot associate a perceived face with a name or set of properties belonging to a specific individual, such as one's husband or hairdresser. Neuroanatomically, prosopagnosia has been associated with lesions of the right fusiform gyrus and surrounding tissue (Wada & Yamamoto, 2001; Sorger et al., 2007), which brings up the question of how cell groups in this region organize different recognition abilities that can be impaired in distinct types of agnosia (e.g., object-shape recognition, face recognition, topographic recognition).

Taking the visual agnosias—including prosopagnosia as a domain-specific agnosia with a strong memory component—together with akinetopsia and achromatopsia makes us appreciate how specific the loss of particular aspects of visual consciousness can be. Whereas massive lesioning of areas V1–V2, the retina, or LGN leads to complete perceptual blindness, visual awareness can in principle be dissected in a piecemeal fashion, taking out V5/MT, V4, and different functional subsets of the fusiform gyrus. Perception of individual object features such as lines and curves may be preserved, but integration into larger composites can be lost as in integrative agnosia. Or, as we have seen, this integrative stage might still be intact, whereas the linkage to a memory trace of a familiar face may be lacking. Above I have argued that densely amnesic patients such as H.M. are (undoubtedly) conscious, but prosopagnosia illustrates how consciousness is impoverished when the linkage between perception and memory is severed.

Having seen how local brain damage knocks out very specific attributes of conscious experience, one may wonder whether there might be a single brain region with an overarching function, integrating the whole gamut of attributes. Such a single “master” region is unlikely to exist although the role of the posterior parietal cortex in spatial vision and hemineglect should be mentioned here. Lesions in the inferior (ventral) parts of the right parietal cortex render patients incapable of attending and responding to objects presented in the left (contralateral) visual field, whereas perception of objects in the right visual field is unhampered (figure 3.8; Singh-Curry & Husain, 2009; Dehaene & Changeux, 2011). This leads, for instance, to difficulties in bisecting horizontal lines running through the visual field. In conjunction with the inability to access visual information for manipulation and attentional selection, hemineglect presents a loss of awareness of the contraleisional visual space (Karnath et al., 2001; Driver & Mattingley, 1998). Stimuli in the neglected hemifield may still be unconsciously processed to support object identification, passing under the “radar” of visual awareness. In addition to the posterior parietal lobe, other regions may be critically involved, most notably the superior temporal cortex of the right hemisphere (Karnath et al., 2001) and dorsolateral premotor areas (Vallar, 1998).

Our overview of visual perception raises the question of whether other modalities are organized in a similar way. About 95% of the literature on consciousness pertains to vision, and it is obvious that our knowledge of other modalities has been lagging behind. Olfaction and taste are intriguing modalities because the chemical senses differ from vision in

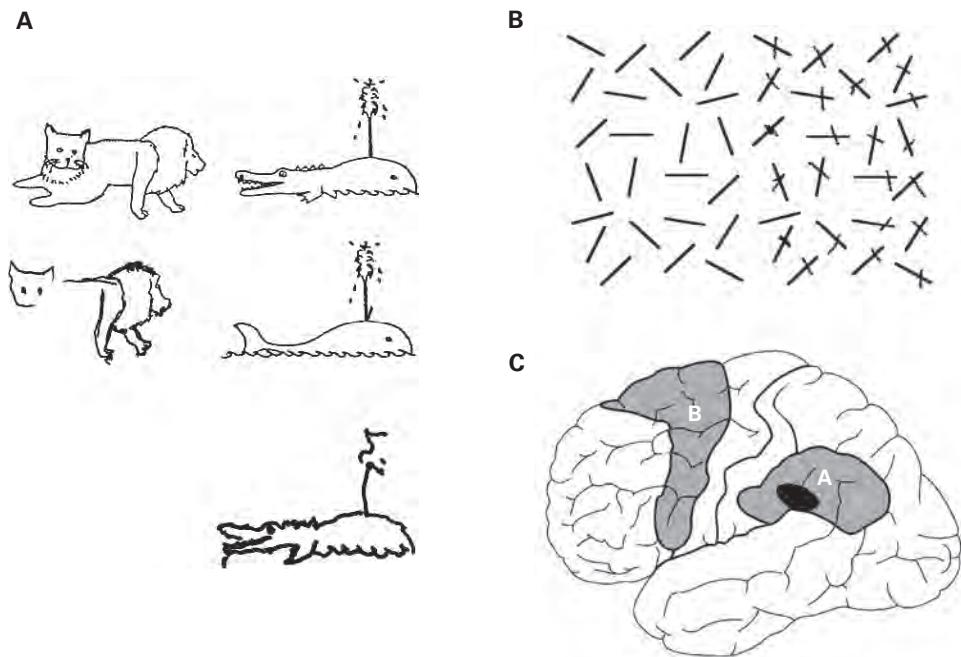


Figure 3.8

Hemineglect in patients suffering from damage to their right hemisphere. (A) While copying figures, patients with hemineglect typically leave out details from the left side of the original. The upper left figure shows the original drawing of a cat-lion chimera, while the patient's drawing is shown below. When facing a chimera of a crocodile and a whale (right-hand side), another patient judged that both sides of the drawing represented whales. This particular patient was able to trace the contours of the two drawings (crocodile-whale and whale) on some trials (bottom right) yet failed to identify the animals implicated here correctly. The dissociation between visuomotor tracing and deficient identification indicates the complexity of hemineglect disorders. (B) Results from a line cancellation task in a hemineglect patient. The patient is asked to mark each of the line segments presented on a paper in front of him yet fails to complete this task for lines on the contralateral side (i.e., in the left visual field given a lesion in the right hemisphere). (C) Anatomical areas implicated in hemineglect. In many patients lesions have been found in the inferior parietal cortex (the dark gray area "A"), around the temporoparietal junction, where the temporal lobe meets the parietal cortex (shown in black). Lesions of the dorsolateral premotor areas (B) and medial frontal regions have also been implicated in hemineglect, although less frequently. Although spatial hemineglect is predominantly associated with damage in the right hemisphere, this scheme projects lesions onto the left hemisphere. See the text for further possible anatomical correlates. (A) and (C) from Vallar (1998); (B) from Driver and Vuilleumier (2001); (A-C) reproduced with permission from Elsevier.

so many ways. The olfactory analogue of blindness is *anosmia*. This loss of conscious olfaction can come in a general form (complete loss) or specific form (lost sensitivity to a specific odorant). As olfactory information is propagated all the way from the nasal epithelium, via the olfactory bulb, piriform cortex, and olfactory tubercle to the thalamus and frontal cortex, there are many way stations where damage could result in loss of smell. Nonetheless, as for vision it is also the neocortex—especially the right orbitofrontal cortex—that is most clearly associated with conscious odor sensations (Gottfried, 2010; Li et al., 2010). Epileptic activity involving orbitofrontal regions sometimes results in cacosmia, the hallucinatory perception of repugnant smells.

Similarly, taste signaling depends on a long chain of lower sensory stations originating at the taste buds in the tongue, but taste perception is predominantly associated with the gustatory cortex, anatomically corresponding to portions of the insular cortex (figure 2.10; Small, 2010). Loss of taste perception—*ageusia*—is caused by lesions of the gustatory cortex but may also occur when the corresponding relay nucleus of the thalamus is damaged (Kim et al., 2007). Whereas the perceptual intermingling of odor and taste during a good meal is universally recognized, little is known about how this integration comes about at a thalamocortical level (Dalton et al., 2000).

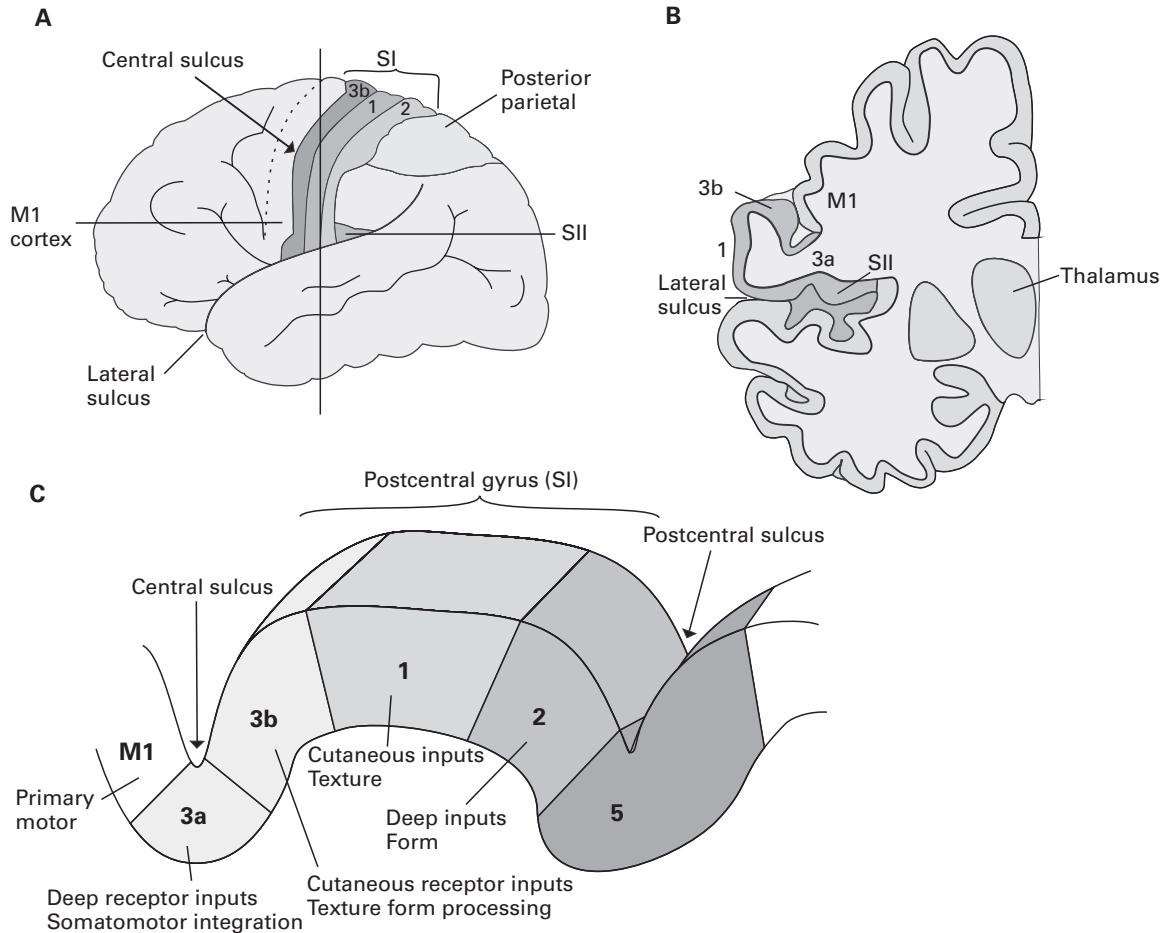
When comparing *hearing* to vision, one might expect that attributes of auditory percepts—such as pitch, source location, or higher-order features such as consonance, dissonance, and timbre—can be coupled to specialized auditory regions in the brain, but this is only partially so at the current time. The primary auditory cortex (A1), with its tonotopic map (see figure 2.11B), is surrounded by secondary auditory cortex, comprising a number of “belt” (and parabelt) areas with their own maps, but we understand yet little about how such different attributes or submodalities are elaborated and analyzed in these areas of the primate brain (Eggermont, 2001; Recanzone & Sutter, 2008). Already in A1, neuronal responses appear to transcend the level of simple feature detection and are more sensitive to combinations of sound components, suggesting a more advanced level of analysis than found in V1 (King & Nelken, 2009). Particularly in nonprimate vertebrates, evidence has been found for a specialization of auditory areas for different attributes. Mustached bats, for instance, hunt for insects with a biosonar system, and their brain is fit with different areas analyzing diverging properties of sonar pulse perception. A portion of the bat’s cerebral cortex is concerned with the global frequency of sonar pulses, whereas another area tunes in on a fine range of Doppler-shifted frequencies which are reflected back by prey toward the ear, informing the system about approaching or receding motion of targets. Yet other areas are specialized for delays between emitted and received sonar pulses, allowing the computation of distance to target (Suga, 1989; Recanzone & Sutter, 2008). In humans, comprehension of speech associated with Wernicke’s area can be mentioned as a specialized form of auditory analysis that rivals in complexity the higher visual system. When the auditory cortex is viewed together with the entire temporal lobe and parietal cortex,

a hierarchy for processing auditory objects and scenes has been recognized, supported by various forms of auditory agnosia (Goll et al., 2010).

The sophistication of our *somatosensory system*, with its wealth of subcutaneous mechanoreceptors, is easily underestimated. With exotic names like Ruffini's and Pacinian corpuscles and Merkel's disks, this is one of my favorite sensory systems, reminiscent of regional Italian recipes and German *Kartoffel* dishes. When our fingers touch an object such as a key hidden in a pocket, different features are signaled by various mechanoreceptors, some being more sensitive to fast or slow changes in pressure on the skin, others signaling static pressure. Via relays in the spinal cord and medulla, tactile information is transmitted to the ventral posterior thalamus and hence to the primary somatosensory cortex (see figure 2.11C) as well as to a portion of insular cortex. As soon as we start actively exploring the key by moving our fingers, changes in muscle tension will be signaled by proprioceptors which aid in identifying the shape, size, and surface texture of the object. Stereognosis—perception of three-dimensional shape through touch—seems to be such a trivial and self-evident process that we would almost forget that the sensation of a coherent object can only arise by the integration of dispersed information derived from spatially distributed mechanoreceptors. The necessity to “bind together” feature information signaled by distributed sources is thus not limited to vision.

Does the somatosensory system also harbor “nodes” coding for specific attributes of touch perception, as we saw for vision? Somatosensory submodalities include light touch, pain, thermal sensations, vibration sense, weight, texture, shape, and joint position.

Figure 3.9 shows how the various subareas of the somatosensory cortex are positioned and connect to each other. Brodmann divided this cortex into area 3a, 3b, 1, and 2, and the most “primary” regions of these are area 3a and 3b, as they are the principal targets of thalamic afferents. Many inputs to areas 1 and 2 come from area 3a and 3b, and all four areas propagate information to higher-order areas such as secondary somatosensory cortex (SII) and parietal area 5. Lesions of area 3b primarily cause general deficits in touch, including poor discrimination of texture, size, solidity, and object shape, as would be expected from a primary cortical station that is somewhat comparable to area V1 in the visual domain. Lesions of area 1 predominantly impair texture discrimination whereas damage to area 2 impairs size and shape distinctions (Randolph & Semmes, 1974; Darian-Smith et al., 1982). Following the processing stream to the secondary somatosensory cortex, neurons become more specialized in macrogeometrical features of objects, integration of stimulus features sensed across disparate skin locations, and are more clearly linked to conscious feeling (de Lafuente & Romo, 2006; Fitzgerald et al., 2006; Meyer, 2011). Altogether, current evidence suggests that “essential nodes” for specific submodalities may be also present in our tactile system. The analogies to the visual system go further, as somatosensory forms of hemineglect (“hemianesthesia”) and agnosia have been described (Vallar, Bottini, et al., 1991; Kilgour et al., 2004).

**Figure 3.9**

Overview of cortical areas involved in primary somatosensory processing in the human brain. (A) Lateral view of human brain showing the location of primary somatosensory areas in the postcentral gyrus. The central sulcus has been partially opened to expose primary motor cortex (M1) and somatosensory area 3b, 1, and 2 (area 3a lies hidden beneath the sulcus; see B and C). Collectively, Brodmann's areas 3a, 3b, 1, and 2 form the primary somatosensory cortex (SI). SII: secondary somatosensory cortex. Higher-order areas involved in somatosensation, such as insular cortex, are not shown here. Based on Keysers et al. (2010). (B) Coronal section of human brain showing primary somatosensory areas and SII. The anterior-posterior level of this section is indicated in part A (vertical solid line). Adapted from Kandel et al. (2000). (C) Divisions of the postcentral gyrus in the human brain, with numbers referring to Brodmann areas (note that secondary somatosensory cortex is not shown in this scheme). Functional processing properties, such as texture form processing, are based on studies in nonhuman primates. Based on James et al. (2007).

3.10 Revisiting Functionalism

What did we learn from Harry's reckless experiment? We witnessed how his conscious life could be stripped of different kinds of memory, emotions, motor capacities, and even language. Certainly our conscious lives would be severely impoverished if we were to be deprived of these abilities, but we would still be conscious. The converse case of intact motor behavior in the absence of consciousness underscored the dissociability of these two domains. Only when specific sensory brain areas are taken out do we see a selective breakdown of consciousness. Lesions of primary sensory areas of neocortex deprive us of entire modalities, while damage to specific higher-order nodes selectively removes motion vision, color, or the tactile feeling of texture.

Classic functionalism posits that a mental state acts as a causal intermediate between sensory input and motor output, but neurological evidence indicates—rather overwhelmingly—that conscious states persist in the temporary or long-lasting absence of motor behavior, including language. Instead, I propose that conscious perception turns out to be an end point of sensory processing (or, alternatively, of imagery), a state that can exist independently of the acute presence of motor or linguistic activity elsewhere in the central nervous system. This does not imply that motor activity would be irrelevant for the ontogeny of consciousness. Coming back to Dennett's characterization of consciousness as a “discriminative state” that disposes creatures to behaviorally react to inputs, I surely would agree that conscious perception must depend on some kind of discriminative state in the brain (otherwise we could not perceive anything specifically), but the weight of evidence shows that consciousness implies more than that neurons will simply generate stimulus-discriminating response patterns biasing the organism toward specific actions. Achromatopsia, for instance, warrants the assumption that, in area V4 and connected areas, special operations must be going on that are not present in the same form either at lower visual or frontal levels.

Revisiting the “hard” and the “easy” problems of consciousness research, our overview lends dual support for maintaining this distinction. First, the “easy” problems of memory, emotions, volition, and motor behavior, and so forth, are by current standards more empirically tractable than the “hard” problem of qualia, as Chalmers (1996) argued. We already know bits and pieces of the neural mechanisms mediating emotions, memory operations, and planned behavior, and what we do not know is often open to investigation with known methods. Secondly, consciousness does, in principle, survive the removal of memory, emotional, and motor capacities but cannot do without those sensory systems that appear to be key for qualitatively differentiated perception. Thus, memory, emotional, and motor capacities are argued to be peripheral to consciousness, whereas perception forms the core of it.

4

What Neural Network Models Can and Cannot Explain about Cognition

It is admitted that neither bodies nor the images of bodies enter the brain. It is indeed impossible to believe that colour can be conveyed along a nerve; or the vibration in which we suppose sound to consist can be retained in the brain: but we can conceive, and have reason to believe, that an impression is made upon the organs of the outward senses when we see, hear or taste.
—Charles Bell (1811)¹

4.1 Construction and Operation of Neural Networks

Shifting our viewpoint away from “wet” brain tissue and neurology, we will now see how mathematical models of neurons and networks have contributed to understanding brain-cognition relationships. Paralleling the explosive growth of experimental neuroscience since the 1970s, the development of neural network models marked a revolution in cognitive science and theoretical neuroscience, but this area of research was initiated much earlier, going back to the work of Warren McCulloch and Walter Pitts at the University of Chicago in the early 1940s. They thought of neurons as computing elements, or building blocks for networks that can solve cognitive problems, at a time when the basic functioning of neurons had already been described although the membrane mechanisms underlying action potential generation were still poorly understood (McCulloch & Pitts, 1943). McCulloch and Pitts’s claim holds that networks of simplified model neurons are capable of universal computation: they can, with appropriate connectivity, solve the same computational problems that an ordinary digital computer can address as well. They can do so in a *distributed* way, meaning that there is no central neuron or computational unit that acts like the central processor in a personal computer. The computing is done by the action of many interconnected neurons together. Cognitive or computational function is not accomplished by an individual neuron but rather by a group of cooperating neurons. Considering that Donald Hebb’s seminal theory of cell assemblies, including his postulate on learning and neural plasticity (see chapter 2), would be published 6 years later (in 1949), McCulloch and Pitts’s work was revolutionary indeed. As the title of their paper (“A Logical Calculus of the Ideas Immanent in Nervous Activity”) already tells us, they did not follow the tradi-

tional, behaviorist idea of nervous activity as just a relay between sensory inputs and motor outputs, but as a way to compute *ideas*. They allocated these ideas not to some abstract, Platonic realm or Cartesian mind space but stated that ideas were *immanent* in nervous activity, implying they are inseparable from it.

McCulloch and Pitts designed a model based on the physiology of the neuron as known at that time. Chapter 2 presented a scheme of neurons with dendrites as elements that receive synaptic inputs, somata as elements converting inputs into all-or-none output (spikes), and axons and their terminals as means to relay the cell's output to target neurons. Crucially, McCulloch and Pitts ignored a wealth of details, such as the complex spatial geometry of dendrites, and produced a reduced version of the neuron that would capture its basic *functional* operation. Apt for their mathematical approach, the minimalist architect Ludwig Mies van der Rohe adopted the *adagium* “Less is more” around the same time.

In the McCulloch–Pitts neuron shown in figure 4.1, the dendrites are anatomically continuous with the soma, giving rise to an axon connecting to target neurons (not shown). Assuming the synaptic inputs are excitatory and the cell is at rest (i.e., nonspiking) when inputs are absent, the activation of a single synapse will depolarize the neuron, mathematically captured by a positive-valued input. The amount of depolarization depends on two factors. The first is the strength of the synapse, which is a function of the quantity of transmitter molecules released, the number of postsynaptic receptor channels, the extent to which the channels open up, and so on. The second factor is the intensity of the presynaptic cell’s activity. If the synapse is very potent but the presynaptic cell is inactive, the net impact on the receiving neuron is zero. The impact of the presynaptic cell on its target cell can thus be captured by the product of presynaptic activity (a_j) and synaptic strength (or

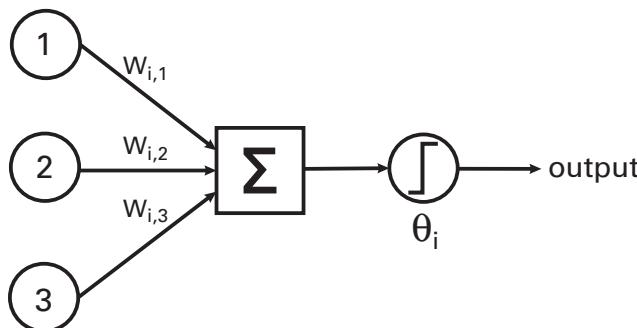


Figure 4.1

Scheme of the McCulloch–Pitts neuron. In this model, the neuron is reduced to its computational core elements, comprising a dendritic compartment that summates (Σ) incoming synaptic inputs from presynaptic neurons 1, 2, and 3 and a somatic compartment that performs a threshold operation (θ_i) and emits spike trains as outputs to target neurons. The subscript i denotes the identity of the postsynaptic neuron, which is usually part of a larger array in a network. The synaptic weight from presynaptic cell 1 to the postsynaptic neuron is indicated by $w_{i,1}$.

weight, w_{ij}), where the subscript “ j ” denotes one of the complete array of presynaptic cells and “ i ” denotes the postsynaptic cell:

$$\text{presynaptic cell's effect} = a_j w_{ij}. \quad (4.1)$$

In real brains, a single synaptic input is usually not potent enough to make a postsynaptic cell fire. At rest, many cells are quietly biding their time at a relatively hyperpolarized membrane voltage, whereas a spike can only be triggered when a depolarization drives the cell up to about -35 mV. Most single-cell synaptic inputs contribute tiny depolarizations in the order of $+1$ mV (Brown et al., 1979). How single synaptic inputs are precisely integrated to a net result that is sensed at the soma and axon hillock is a complex question, which McCulloch and Pitts cleverly circumnavigated by positing a summation rule:

$$\text{Total effect of presynaptic cells} = \sum a_j w_{ij}, \quad (4.2)$$

where the Σ sign indicates that all of the presynaptic inputs are summed to compute the total depolarizing effect on the target cell. In figure 4.1, this summation is equal to $(a_1 w_{i1} + a_2 w_{i2} + a_3 w_{i3})$. Whether or not the total depolarization is strong enough to trigger a spike depends on several physiological factors, but again McCulloch and Pitts made a simplifying assumption. They adopted a fixed spike threshold, which corresponds to a particular membrane voltage. This level has to be exceeded by the synaptic inputs if a spike is to be generated:

$$\begin{aligned} \text{postsynaptic cell's output} &= 1 & \text{if } \sum a_j w_{ij} \geq \text{threshold} \\ && (4.3) \end{aligned}$$

$$\text{postsynaptic cell's output} = 0 \quad \text{if } \sum a_j w_{ij} < \text{threshold}$$

where a value of 1 symbolizes a spike and 0 its absence. This output value applies to a certain window of time. In real life, neurons operate on an awfully fast time scale, and some types can sustain a firing rate of hundreds of spikes per second. In computer simulations, we can follow McCulloch and Pitts and choose to work with model cells that generate an all-or-none output (i.e., “binary” neurons that are either active or inactive but have no state in between). In that case, it is appropriate to frame the neuron’s activity in a narrow time window (e.g., 5 ms) in which we can safely say that a cell fires or does not fire, and the chance of having a third state (e.g., two spikes) is negligible. Alternatively, we may work with model cells whose output is *continuous* between a resting (nonspiking) state and the maximal activity a cell can reach. In this scenario, time windows will be much wider, generally in the range of seconds. Whether we choose an all-or-none or continuous type of neuron is important but not crucial for the way neural networks operate.

Thanks to the seminal work of McCulloch and Pitts, an abundance of different neural architectures was developed. As in the animal kingdom, neural nets can be classified into different species and families, based on their architecture, the type of cognitive task the

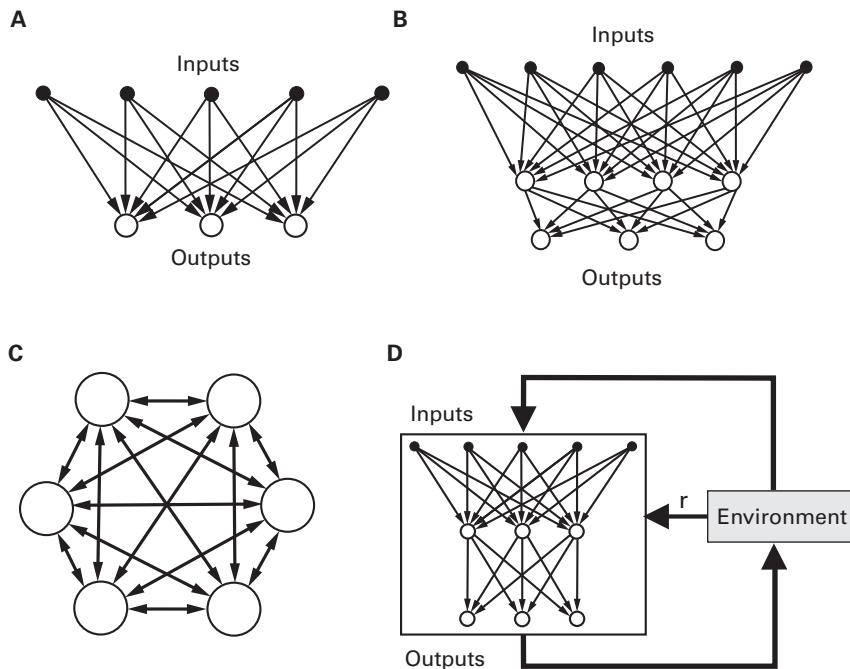


Figure 4.2

Different neural network architectures. (A) A one-layer feedforward network or simple perceptron is illustrated. Black dots in the upper row indicate elements relaying inputs to the lower layer. Only the synapses (arrows) onto the neurons in this lower layer are modifiable. (B) A two-layer feedforward network is depicted. This is an extension of the network in (A); the intermediate layer is often called a “hidden layer.” (C) A recurrent network is illustrated. Characteristic of this type of network is that neurons project back on neurons or neural groups from which they receive inputs. In this case, the bidirectional arrows symbolize bidirectional connections, and all neurons are connected to each other. External inputs and efferent projections are not shown here. (D) The basic architecture for reinforcement learning is depicted. The central, two-layer feedforward network is an example of a network that may be trained on this kind of algorithm. In addition to sensory inputs that instruct a trained network to produce adaptive motor outputs, the environment also provides a reinforcing feedback (r) that conveys a scalar signal to the network, on the basis of which its internal connections can be modified. (D) adapted from Hertz et al. (1991).

network should perform, or the procedure to train a net on the task. Main classical architectures comprise feedforward neural networks, recurrent networks, and hybrids of these two forms (figure 4.2; Hertz et al., 1991, Churchland & Sejnowski, 1992).

The most basic feedforward network, called a *one-layer* (or *simple*) *perceptron*, consists of a layer of neurons receiving presynaptic inputs (see figure 4.2A; Rosenblatt, 1961). Every presynaptic unit² connects to all neurons in the layer, and the synapses are modifiable, meaning that their weight (w_{ij}) can be changed. Synaptic changes do not occur haphazardly but are regulated by *learning rules*—for instance, based on the way that induction of LTP or LTD depends on pre- and postsynaptic activity (see chapter 2). In a perceptron,

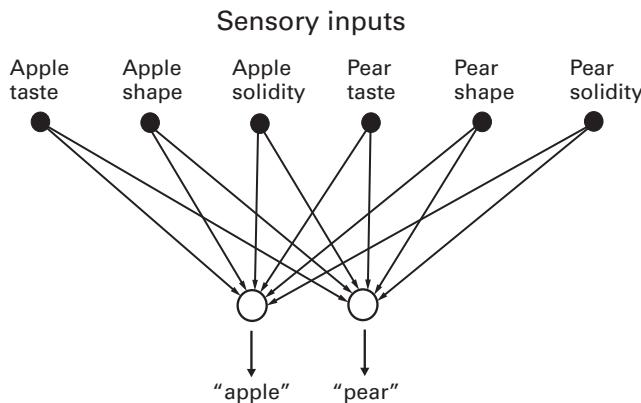
presynaptic units just supply the neurons in the (postsynaptic) layer with inputs and do not contribute to any computation, whereas the neurons in the layer sum up the synaptic inputs and will individually fire if their total excitatory input is large and remain silent if it is too low (see equation 4.3).

Following Rosenblatt, Minsky and Papert (1969) showed that some essential problems such as the Exclusive-OR (XOR) problem could not be solved by a one-layer perceptron. The XOR problem holds that neurons in the postsynaptic layer receive synaptic inputs from two sources (x and y) and have to generate activity ($a_i = 1$) when either one of these sources is active, but not both at the same time. If the sources have the same activity status, then the neurons should remain silent ($a_i = 0$). To enable feedforward networks to solve this problem, an intermediate (or hidden) layer needs to be inserted in between the input neurons and the output layer (see figure 4.2B). A perceptron including one or more hidden layers is also called a multilayer feedforward network.

In contrast to this purely feedforward flow, a *recurrent network* is characterized by a group of neurons whose output is fed back to themselves (figure 4.2C) or to a second group of neurons projecting back to the first group. Whereas a feedforward network is merely “reactive” in that its output is a function of its inputs, a recurrent network allows the output of high-level neurons (i.e., close to the net’s output) to act as input to lower-level neurons (i.e., close to its input), offering different computational possibilities. A special case of recurrency is when all neurons are connected reciprocally with one another, which is found in the *Hopfield network* (see below). However, let us first continue with feedforward nets and see how they work.

It goes beyond the scope of this book to explain the functioning of neural networks in detail. What is stunning, though, is how many tasks these feedforward nets can perform given their structure. If the activity of the inputs represents, for instance, taste, visual shape, and solidity (see figure 4.3), then the activity of the neurons in the layer is able to classify the input patterns into categories distinguishable by these three properties (e.g., apples and pears). This pattern classification can also be cast as a recognition process. If the left-hand output unit in figure 4.3, for instance, is active only when apple-like properties are presented, its activity can be taken as a recognition of the input pattern as an apple, whereas the other output unit signals recognition of a pear. Multilayer feedforward networks have been applied to a host of complex tasks, varying from autonomously driving a car (Pomerleau et al., 1991) and flying aircraft, to playing games such as backgammon and chess (Tesauro & Sejnowski, 1989) and making a medical diagnosis. Additional biological applications include identification of faces and recognition of speech or handwritten text (Nowlan, 1990; Müller et al. 2013).

Let us take a closer look at how a net can be trained to distinguish apples from pears. We have many criteria to tell these classes of fruit apart, but the three features used in figure 4.3 provide useful characteristics. Even though different brain structures process these and many more (sub)modalities, we just represent the three types of sensory information by

**Figure 4.3**

Simple perceptron to categorize, or recognize, apples or pears from three dimensions of sensory input. Black dots in upper row relay sensory information on taste, visual shape, and tactile solidity to the two classifier neurons in the bottom layer. If the three left-most units have sufficiently strong and positive connections to the neuron conveying the “apple” output, this classification will be generated when one or multiple apple-like properties are detected, at least more strongly than any pear-like properties relayed by the three right-most upper units at the same time. The same holds for the right-hand “pear” unit, which by training comes to be strongly activated by the three right-most units.

three different inputs for each output category (apple or pear; see figure 4.3). Taking a bite from an apple activates a host of sensory subsystems—the proprioceptors of our tongue and jaw muscles as much as the tongue’s mechanoreceptors and taste buds—but not all of these inputs need to be used here. The three left-most inputs in figure 4.3 represent typical apple-like properties, and the left-most of these represents apple taste. Thus, an activity value of this unit of 1 will represent an apple-like taste whereas a 0 would correspond to absence of that taste and so on. The three right-most inputs represent pear-like features. The whole array of sensory inputs has the nature of a vector, which is a series of numerical values (as opposed to a scalar, which is represented as a single numerical value). For instance, the string [0, 1, 1, 1, 0, 0] can be taken to represent a pear-like taste, an apple-like shape, and a firm bite. Each of these properties is coded by two units, usually assuming opposite values in this example.

By way of the downward axons emitted by each sensory unit, the input vector is now propagated to the postsynaptic layer, and each input fans out toward the neurons in that layer. The axons emanating from an active input unit convey its spike activity to its synaptic terminals, and the connection strengths determine how much impact this activity has on the postsynaptic neuron in the layer (see equation 4.1). The trick to solving the classification problem is to train it so that the input vectors result in the appropriate output patterns. This training is done by presenting many different examples of “apples” and “pears” to the net and modifying the connection weights so that apple-like input vectors more and more result in activation of the left-hand output neuron (where 1 means that the answer is

“apple”) and silencing of the right-hand neuron (where 1 would mean “pear” if this neuron were to be activated). The mathematical operation is captured by the following:

input vector $[0, 1, 1, 1, 0, 0] \rightarrow$ output vector $[1, 0]$

In our apples-and-pears example, this would mean that if the output $[1, 0]$ were correct, the classification of the inputs would be swayed toward an “apple” answer despite the pear-like taste signaled by the 0 and 1 values in the first and fourth input. We can think of this operation as a mapping from an input to an output pattern.

Pondering ways to set the connection weights right, activity in all three left-most inputs should bias the output pattern toward activity of the left output and zero activity of the right one. We achieve this result when the weights from the three left-most inputs onto the left output neuron are all strengthened and those onto the right output neuron weakened (or even made negative, inhibitory). The same treatment for the right half of the net will take care of pear-like inputs activating the neuron outputting “pear.”

The many learning rules that can be used to govern the synaptic weight changes can be divided in two main kinds: *unsupervised* or *supervised* (Hertz et al., 1991). Supervised learning means that there is an entity (or agent) that has a priori knowledge of whether the net’s output is correct or not, or is otherwise able to provide feedback to the net on how good or bad its performance was. The entity can be internal to a larger network system or could be external to it. It may assume extremely varying appearances such as a mathematical look-up table telling exactly which input–output mappings are wrong or right, or the ecological environment an animal is in, providing its brain feedback about the success of its hunting or escape actions. The feedback can come in the form of very detailed, high-dimensional information telling what the correct output of each postsynaptic neuron should have been, so that errors in the actual outputs can be computed and synaptic weights adjusted based on those errors. Such detailed feedback lies at the heart of the back-propagation algorithm (Bryson & Ho, 1969; Werbos, 1974; Rumelhart et al., 1986) and leads to fast learning even in large networks, but it is less plausible biologically because no such virtually “omniscient” entity is known to exist in the brain.

A less informative type of feedback is used in *reinforcement learning* (figure 4.2D). Here the feedback comes as “weak” supervision: it merely issues signals to the network indicating how good or bad the overall network output was (relative to achieving a particular goal, such as food or keeping your balance; Sutton & Barto, 1998). These reinforcement signals affect synaptic strengths such that a given input pattern will have a higher probability of activating the “good” (correct) output pattern on future learning trials. Because the signals are computed as a function of the network’s output but feed back to all of its layers containing modifiable synapses, a prototypical reinforcement learning model presents a hybrid between a feedforward and a recurrent network. The changes in the synapses mediating sensory-to-motor feedforward propagation are governed by a recurrent feedback loop, acting as a “critic” and signaling whether the overall output was appropriate or not.

Slow as learning in these models may be, they are more compatible with neurobiology, as the entity providing the feedback is the animal's environment, in which actions can indeed be classified as good or bad relative to survival and reproduction. Moreover, particularly effective variants of reinforcement learning have been developed (Widrow et al., 1973; Sutton & Barto, 1998), and these bear a strong formal resemblance to the functioning of neural systems involved in similar types of learning, such as mesencephalic dopamine neurons (Schultz et al., 1997) and prefrontal–striatal systems (Pennartz, 1996; Daw et al., 2005; Pennartz, Ito, et al., 2011).

In unsupervised learning, no surveillance agent or “teacher” is around to keep learning processes in check. Yet, unsupervised nets can learn too, and they do so based on statistical relationships inherent in input patterns. An example relationship would be that two elements of an input pattern are strongly coactive (positively correlated), even when other elements of the input vary. When many exemplars of a person's face are presented, many features such as the viewing angle, shading, contrast, and size will vary, but there will also be definable invariants that render the face constant under different viewing conditions. The shape of someone's mouth, nose length, and distance between the eyes can be such defining features. When these features are repeatedly presented to a model, invariant correlations in the inputs will lead to coactivation of synapses converging on the same postsynaptic neuron, boosting its activity. When the model uses a Hebbian learning rule (see chapter 2), it will be able to perform cognitive tasks such as face recognition (e.g., Dayan, 2006) and determining the shortest route along a series of cities spread out across a two-dimensional plane (the Traveling Salesman problem; Hertz et al., 1991; Aras et al., 1999). In addition, unsupervised learning has been applied to problems of *self-organization*, for instance how neurons in the developing visual cortex become selective for the orientation of stimuli in the visual field (Bienenstock et al., 1982). Another application lies in recurrent associative memory networks (see below).

4.2 Advantages of Parallel Distributed Processing

Across the second half of the twentieth century, neural network models have revolutionized the way we envisage links between neuroscience and psychology, between the functioning of single cells and the generation of complex behavioral and cognitive phenomena. Had they not been invented, we might have been conceptually stuck in metaphors of the brain derived from classical Von Neumann–type computers, on which our present day laptops and personal computers are based. Here, information streams needed for a computation at hand converge to a central processing unit (CPU). The CPU's work is dictated by a sequence of preprogrammed instructions, together forming a program. Typically, information is not stored within the CPU but elsewhere (e.g., on hard disk), and the storage and retrieval of information occurs with the help of a label or “address” (a string of symbols or

numbers specifying the storage location). Once a segment of hard disk has been damaged, the information stored at that location is completely lost.

By contrast, the brain is not organized around one CPU but is characterized by massive *parallel distributed processing* (PDP). This is because information is processed by neurons and synapses operating not in series but parallel to each other, each contributing bits and pieces to the overall computational task. This is already evident as light strikes the eye: millions of photoreceptors change their activity state as soon as the eye is opened, resulting in a cascade of activity from inner to outer retina, and activating millions of ganglion cells sending their output to the LGN and other subcortical stations. Having passed the bottleneck of the LGN, the visual stream does not converge to a central, Cartesian processor but fans out across the primary visual cortex. The Felleman–van Essen map of the visual system (figure 3.6, plate 1; see also figure 6.3, plate 3) illustrates how the output from V1 diverges into partially segregate routes for visual analysis, underscoring the point of parallel processing. Concurrently, the brain processes sensory inputs from other modalities—we feel, hear, and smell things at the same time we can see them—not to mention the functions above and beyond perception exerted simultaneously: attention, memory encoding and retrieval, planning, motor behavior, and language.

Simplified and reduced as many neural nets may be, they do capture this fundamental PDP feature of brain organization. When a network has to classify a facial image as male or female, there is no single neuron or axons coding the whole feature set characteristic for the image. Different neurons may code various visual-field segments of the image or alternatively process different global features such as elongation or angularity.

Information is not only *processed* in parallel by neural networks but it is also *stored* in distributed form. There is no single site or connection in the net where the “engram” is located (cf. Lashley, 1943). If a reinforcement-learning network is presented with fruity features and trained to distinguish apples from pears, the reinforcing feedback reaches all modifiable synapses in the network, and as a result many of them will be modified, thus storing the classification knowledge in distributed form. Hence, neural network modeling has greatly influenced the view—now prevailing in the neurobiology of learning and memory—that memories are stored by widespread synaptic modifications throughout neuronal networks, including “micro-LTP” and “micro-LTD” processes across millions or billions of synapses, along with distributed synaptogenesis or synaptic elimination (King et al., 1999; Trachtenberg et al., 2002).

PDP has distinct advantages compared to central processing. As compared to CPUs performing about 100,000,000,000 floating-point operations per second, a single biological neuron is dramatically slow. The membrane time constant of cortical pyramidal cells, indicative of how fast they can integrate synaptic inputs, is in the order of 10–20 ms (Zhang, 2004; Oswald & Reyes, 2008), and its maximal firing rate “only” reaches about 20–100 Hz. However, this sluggishness is compensated for by the parallel-distributed nature of brain organization, as layers of neurons can process multiple features or segments of the

input at the same time without resorting to a central “funnel” through which all information must pass. Think of how our brains, seemingly without effort, process a nearly continuous flow of visual information.

Secondly, neural nets couple speed with high storage capacity. One and the same neural net can store many different visual images or different types of information, such as melodies, movement sequences, or tactile patterns. In contrast to a conventional hard disk, a neural net can store many images in superimposed form (Hopfield, 1982, 1984; Sompolsky & Kanter, 1986). All features of a stored image are embedded in a distributed way, and the sheer number of synapses endows the net with redundant capacity to store many images simultaneously until the network has become saturated.

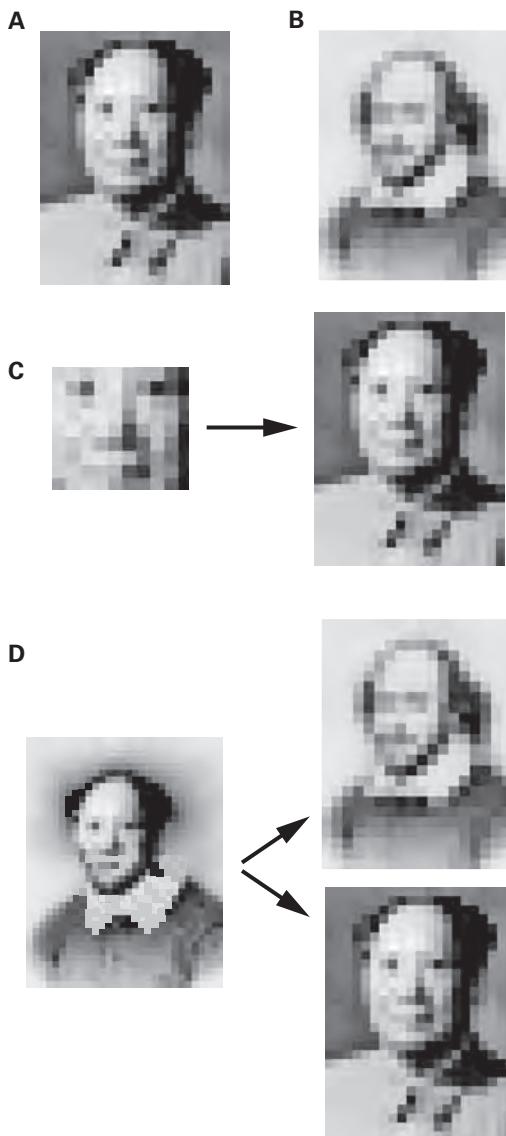
A third advantage is that neural nets do not operate by fixed rules or programming instructions to solve the cognitive problem at hand. If a conventional computer tries to distinguish apples from pears, it will do so based on a series of stereotyped operations, beginning with the calculation of shape features, comparing those features to the standard criteria for apple- and pear-like classes it stores in memory as rigid numerical values. For a neural net, these intermediate instructions need not be programmed, and it has no look-up table for setting criteria on which decisions are based. A well-trained net will directly convert the input vector into an output while the learned settings of the synaptic weights take care that the input space is correctly parceled into subspaces of apple- and pear-like features. Their operation is not hampered by tiny programming errors, which can so easily lead a conventional computer program to crash. If a few neurons in a large enough net become dysfunctional, it will remain well-behaved because of the majority of units that remain healthy. By novel learning, weights can be readjusted so that malfunctioning units will be compensated for.

This flexibility is also expressed when the context in which a cognitive task must be solved is changing. Telling apart apples from pears is a job an organism performs even under changing lighting conditions, including at midday and sunset. Whereas a conventional computer program requires separate instructions to deal with spectral changes of impinging light, a net’s learning rules can make it adapt to such changing environmental circumstances.

4.3 Emergent Properties of Neural Nets

Let us now make a leap from feedforward or hybrid neural networks to a class of recurrent networks that illustrate emergent properties of nets par excellence.

Imagine you are making a trip to Beijing and run into a progovernment march through the city center. Suddenly, amid the crowds, you discern a fuzzy but familiar face in the distance (see figure 4.4A), depicted on a big poster board carried by demonstrators—is it the Great Helmsman? After some hesitation you “know it immediately”: it is Mao, and you stretch out above bystanders to see more of the picture.

**Figure 4.4**

Pattern completion in a recurrent neural network. (A) and (B) show two coarse pixelated patterns of portraits (based on www.opticalillusion.net/optical-illusions/pixelated-faces), where each pixel represents a neuron (gray value corresponding to neural activity, from 0 to 100%). Connections between neurons are not shown. (C) When only a fragment of Mao's face is presented as input to the net, this incomplete pattern will trigger completion of the full face in a number of iterations. Note that only the fragment is shown here, not all surrounding pixels that do not convey information about the identity of Mao's portrait. Here, successful pattern completion assumes no other faces similar to the presented fragment had been stored by previous training. (D) In a network that has been previously trained to store both Mao's and Shakespeare's face, presentation of a hybrid image may lead the network to restore either Mao's or Shakespeare's portrait. This restoration takes multiple iteration steps in network states. Which face is eventually "chosen" depends on several factors such as noise and exact similarity of the presented test pattern to either one of the originals.

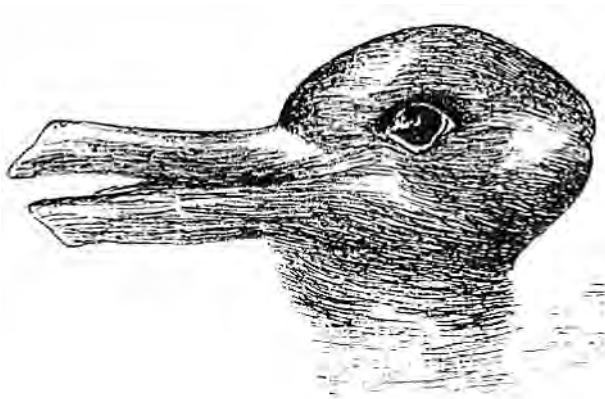


Figure 4.5

Ambiguous duck–rabbit figure introduced to psychology by Jastrow (1900) but sometimes attributed to Ludwig Wittgenstein. This type of figure is also labeled “bistable” as a single visual input pattern can result in two stable (although switchable) perceptual interpretations of the image.

Next to recognizing a face from a fuzzy or incomplete image, the phenomenon of all-or-none interpretation of ambiguous visual images is all too familiar as illustrated by Jastrow’s duck–rabbit in figure 4.5. How can we build a “minimal” neural network capturing this capacity? If we were to resort to feedforward nets, the problem would arise that these can only execute mathematical *functions*; a single input pattern results in a single output (in this case, an image interpretation), whereas here we can seemingly freely choose to alternate between two different interpretations given a single image.

In the early decades of the twentieth century, the Gestalt psychologists Kurt Koffka, Max Wertheimer, and Wolfgang Köhler identified perceptual principles which our brains may use to group elements into objects and segregate “figure” from background. If two visual elements roughly move in the same direction at a similar speed, this so-called “common fate” provides a clue to the visual system that these elements can be seen as part of the same object (see figure 4.6). If we have to segregate a set of intertwined computer cables, a single cable can be recognized and traced based on the collinearity and continuity of its elements, even if its image is intersected by overlapping cables (“law of good continuation”). Grouping of elements may also be based on proximity, similarity, closure, and further principles, but the problem remains unsolved of how various low-level clues converge to a common, overall interpretation of the image, which in our perception can “flip” between a duck and a rabbit.

Work on recurrent network models initiated by Little (1974), Willshaw (1989), Kohonen (1984), and others and brought to fruition by Hopfield (1982, 1984) has shed some light on this problem by showing that fairly “minimal,” autoassociative networks can achieve whole-pattern recognition and classification. Autoassociative networks form a specific

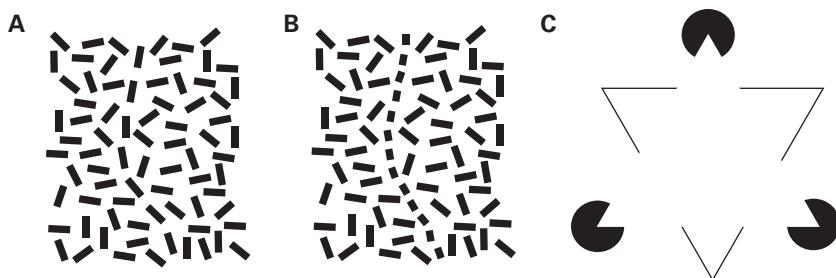


Figure 4.6

Gestalt properties used to group individual scene elements into objects and segregate “figure” from “ground.” (A) Amongst this set of rectangular elements, the common orientation and collinearity of a subset of elements can be used to distinguish a curved line running from top to bottom. If these elements would coherently move in the same direction whereas background elements would move in one or multiple different directions, common fate will act as an additional Gestalt cue for grouping. (B) Same as (A), but now similarity is added as a Gestalt cue to segregate the curved line more readily from the background. (C) The triangle of Gaetano Kanizsa derives much of its illusory power from the Gestalt principles of closure and continuity.

subclass of recurrent architectures (see figure 4.2C) in that the output of the neurons is directed to fellow neurons although they may also receive external input and project to a separate output layer (see figure 4.7). Their core property is that a group of neurons projects back onto itself (hence the adjunct “auto”), so that the output of one neuron comes to be associated with the output of neighboring neurons synapsing on the same target neurons.

Let us take a closer look at how a recurrent network can achieve global pattern recognition. We depart from two prototype patterns—pixelated versions of Mao’s and Shakespeare’s face (figure 4.4)—and first need the network to store these two patterns. This early training can be considered an “imprinting” phase and is marked by abundant synaptic weight changes throughout the network. Basically compatible with the neurophysiology of synaptic plasticity, a Hebbian learning rule can be used for this imprinting. There are many ways to represent a visual image of Mao’s face as a whole, including a coding scheme based on a collection of shape, color, texture, and surface reflection features, but for the simplicity of the argument we assume the image is just coded in black-and-white pixels (Hertz et al., 1991; note, however, that in figure 4.4A gray values are used; cf. Hopfield, 1984). Each pixel represents a neuron, turning “black” when active ($a_i = 1$) and “white” when inactive ($a_i = 0$). The whole image is fed into the neural network, and each pixel of the original image determines the activity state of one corresponding neuron. All neurons are connected to each other, and in the case of the original Hopfield models, the neurons are also reciprocally connected. The Hebbian learning rule takes care of two different scenarios applying to pairs of pre- and postsynaptic neurons that are exposed to an input pattern:

1. If the pre- and postsynaptic cell are both active ($a_i = 1$ and $a_j = 1$), then set the connecting synapse to a strength of +1.

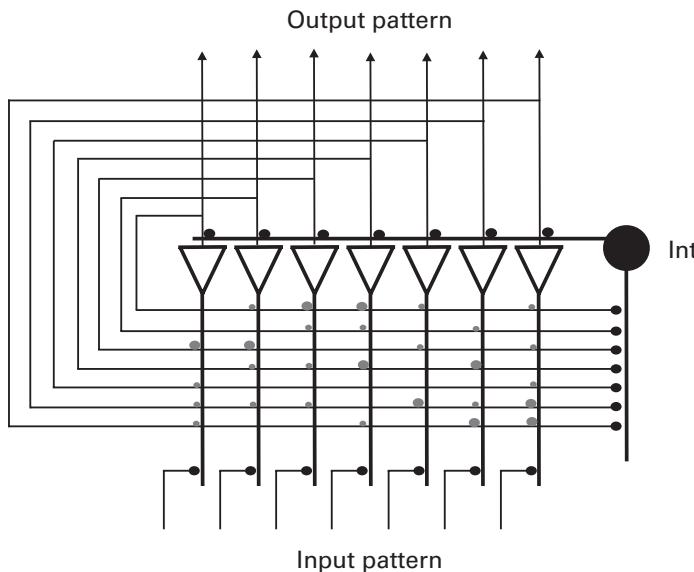


Figure 4.7

Autoassociative recurrent neural network. Input patterns are relayed to the network via synapses on the distal dendrites of pyramidal-cell-like neurons (triangles with descending thick lines). Axons are emitted from the cell bodies, and each cell's axon branches into an output sent to external targets (not shown) and a recurrent collateral that forms modifiable synapses (shown as gray ellipsoid elements) with fellow neurons in the same array. Associative memory is stored in the matrix of recurrent connections. Variable synaptic strength is symbolized by the varying size of the synaptic contacts. The collaterals also synapse onto an Interneuron (Int; black), which controls excitability of the net via separate inhibitory connections onto the cell body or axon hillock. Based on Durstewitz et al. (2000).

2. If the pre- and postsynaptic cell are both inactive ($a_i = 0$ and $a_j = 0$), then set the connecting synapse to +1.
3. If the pre- and postsynaptic cell are anticorrelated in their activity ($a_i = 1$ and $a_j = 0$, or vice versa), then set the connecting synapse to -1.

We can think of the imprinting phase as a sequence of image presentations, where the same image is repeated several times, giving rise to small, incremental weight changes. As a computational shortcut, one can also set the weights directly to their full-blown values in one strike, according to rules 1–3. In both procedures, multiple images can be presented and stored, although there is an upper limit the network may absorb (around 14% of the total number of neurons in the network; Hertz et al., 1991).

Let us see what happens if degraded (incomplete or very noisy) images are presented to a network that has already stored two pixelated but “clean,” unambiguous prototype images of famous people (Mao and Shakespeare; see figure 4.4A and B). This part of the task is referred to as the testing phase. If our recognition network receives only a fragment of Mao’s face (see figure 4.4C), the activity state of a majority of neurons will randomly

float about, but a minority will be activated according to the pixel pattern of the intact part. The active (black) neurons representing this part will activate some neurons that were initially randomly active because the weights of their connections are positive. The same neurons will also turn other neurons in the noisy parts of the image to inactive (white) because during prototype training their activity states were anticorrelated and their connections set negative. The inactive neurons of the intact part play no role in this restoration process. However, in the next cycle of activity (or iteration step), all neurons activated in the previous step will activate more neurons (if their activity was also correlated during training) and inactivate other neurons, being anticorrelated to them in the prototypical Mao image stored earlier. This process results in a completion of the original Mao pattern across several cycles.

The type of associative memory we have construed now is labeled “content addressable” because a complete memory pattern can be retrieved by presenting part of its original content, not by a separate address label as in conventional computers. Content addressability is considered more biologically realistic than the conventional alternative.

Patterns may not be just incomplete or noisy but could also resemble two of the stored prototypes, giving rise to ambiguity. What happens when an image is presented that is a cross between Mao and Shakespeare (see figure 4.4D)? Network simulations show that the collective activity pattern may evolve over consecutive activity cycles toward one of the two prototypes, but more gradually than when an unambiguous resemblance was present. This dynamic evolution over time is also described as convergence or relaxation from an initial to a final state. Analogous to physical lattice structures called spin glasses, Hopfield (1982, 1984) described the evolutionary states of a system by an “energy function.” Specifically, whenever the pre- and postsynaptic activity of a pair is consistent with their connection weight installed by pretraining, the neuron pair is taken to contribute negatively to the net’s energy. But whenever a pair’s activity values are inconsistent with the previously stored pattern, it contributes positively to the energy. By summing up the contributions of all cell pairs in the net, the energy is computed as a collective property applying to the whole network. Metaphorically, a very positive energy across the whole net reflects a state of “tension,” indicating the net is still remote from its final state. In contrast, a strongly negative energy indicates the actual activity state is close to the stored pattern. The convergence from an initial noisy or dissimilar pattern to a stored image can thus be captured as a gradual walk in the energy landscape, going downhill on average (see figure 4.8).

Sometimes, when the test input is equally similar to two prototypes, the network gets stuck in an intermediate, spurious state, requiring another boost of new input to kick it out of this misrepresentation. The associative memory described so far is an example of a network displaying emergent properties because it holds greater complexity and dynamics than it would have if it were merely the sum of its parts. Here, “emergent” denotes a novel behavior or property of a collection of functional elements as a whole, not attributable to a

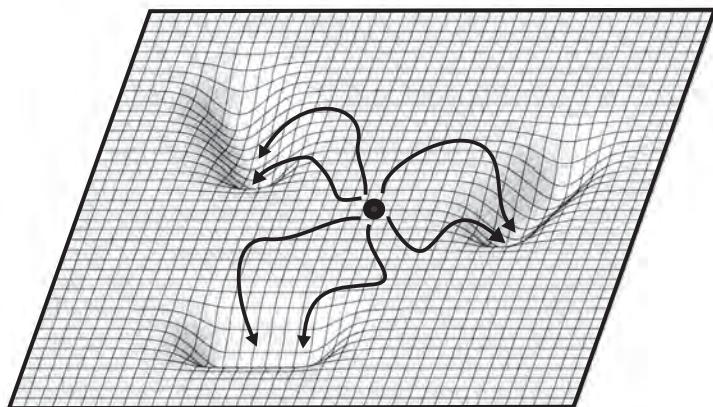


Figure 4.8

Energy landscape of an associative memory network. The X–Y plane represents the possible states of the network and the Z direction the energy of the corresponding state of the network. A multineuron network has many more dimensions than the surface plotted here, as each neuron's firing pattern constitutes a separate dimension. For visibility this high-dimensional state space has been compressed into two dimensions. Neighboring points correspond to similar activity patterns. An initial state of the network is indicated by the position of the ball, and the trajectories of this ball across the surface reflect potential walks of the network through state space toward an activity configuration with low energy. Three basins of attraction are shown, and the network may reach each basin via different trajectories, depending on noise or spontaneous fluctuations. The downhill walk from high to low energy can be described as a process of pattern completion and a transition from “tension” to “relaxation.” Adapted from Morita and Suemitsu (2002); see also Hertz et al. (1991).

single functional element but to the interactions and self-organizing processes in a complex system.

Emergent behavior in recurrent nets appears first and foremost from the fact that, given a noisy or incomplete input, they are able to converge to a stored pattern as a whole—or to a Gestalt, in classical psychological terms. This process of pattern completion works differently than by adding up the results from separate computations on collinearity of pictorial elements or other Gestalt features. The perceptual decision to recognize Mao versus Shakespeare is a collective network effect, instantiated by the convergence of activity patterns toward one of the prototype states. Emergence is also illustrated by a continuing high performance of the network when some of its constituent neurons are lost because information storage is distributed across synapses and the network generally contains so many neurons that pattern information is redundantly encoded. This property of “graceful degradation” points to a key advantage of parallel distributed memory in general, making it more robust and fault tolerant than conventional computer systems. In addition to content addressability, the capacity to handle noisy or incomplete information constitutes another advantage for neural network systems and is also situated at the level of collective neural activity, not directly apparent in single-neuron behavior.

Despite their simplifications, recurrent nets illustrate how a cognitive process such as recognition may unfold in time, by an evolution of network states, eventually converging

to a final state. Such a final state is also described as a stable attractor. In general, an attractor does not have to be stable; it may also consist of a set of subsequent states to which a dynamic system evolves over time. System states that are slightly away from an attractor will converge back to it, making the system tolerant to small perturbations. When attractors are considered in a geometric space comprising all possible states of a system, they may appear as a point (e.g., the fixed spatial position a pendulum assumes when its swing has been damped), a curve (e.g., a limit cycle, such as the fixed oscillation of a pendulum swinging in a vacuum), or even more complex forms, such as strange attractors (Freeman, 1996; Tsodyks, 1999). Much research has been done to examine attractor-like behaviors of neural systems, but for the sake of brevity we will focus on the stable states (fixed-point attractors) that trained Hopfield networks converge to.

4.4 What Do Neural Nets Tell Us about Neural Representations and Consciousness?

What do these emergent network properties tell us about the way the brain generates neural representations, conscious or unconscious? O'Brien and Opie (1999) and others have argued that the brain's capacity to self-organize stability in perception is a crucial hallmark of conscious systems. Under this view, neural networks that can stably maintain explicit representations are argued to be sufficient to explain phenomenal experience. Here, "explicit" refers to the encoding of distinct data items by a physically discrete object, such as a neuron or an allocated subspace on a computer chip. This claim will be examined further, but regardless of the outcome we can at least say that neural nets have shown how pattern stability may be achieved in a unitary, whole-image fashion. This stability might turn out not to be sufficient for generating conscious representation but can be considered one of the hallmarks of steady percepts and may even be one of their necessary properties.

Apart from this remarkable feat, there is another, deeper lesson neural networks have taught us. Their emergent properties tell us how low-level phenomena, such as synaptic modifications, give rise to phenomena at a higher level of organizational complexity. Imagine one of the neurons in the recurrent net of figure 4.7 would be replaced by you, as a person, and how people in your neighborhood would replace the other neurons. This situation slightly resembles an orderly array of dictator-ruled citizens carrying rectangular plates collectively representing their leader and flipping this image toward an audience of presidential party committee members all at once. Once you are placed in this array, you connect to fellows around you by handshakes, and the strength of your handshake represents the connection weight. To connect to more than two persons, you would be endowed with many arms, like the Hindu goddess Durga. Acting as a single neuron, all you would experience are the handshakes you receive from your neighbors, raising your level of muscle activity and prompting you to enhance your grip on your neighbors' hands. However, you would have no clue what you and your neighbors would be representing as

a whole. The cognitive act of pattern recognition should be ascribed to the assembly of agents, not to the individual.

Also Paul Churchland (1995), Patricia Churchland (1986), and cognitive scientists (Baars, 2002) as well as neuroscientists (Crick & Koch, 2003, Koch, 2004; DeHaene & Changeux, 2011) have advocated a neurocomputational account of consciousness. The variant of the Churchlands is “eliminative” in that it attempts to explain all mental phenomena by interactions between the brain’s constituent physical elements; mental phenomena do not actually exist as they are merely constructs used in everyday language and folk psychology. Paul Churchland considered whether a number of salient dimensions of consciousness may be attributed to a neural network having both feedforward pathways and recurrent feedback from the output to earlier layers (see figure 4.2). He argued that the net, resembling a corticothalamic system, may in principle explain dimensions such as short-term memory, steerable attention, and multistability in relation to ambiguous sensory input and the unified nature of experience. Indeed I agree that it is in these cognitive dimensions where neural nets have brought us the most striking advances.

However, there is one nagging “detail” that points to a crucial omission in this account. Most types of network carry out a mapping from an input to an output pattern; mathematically, this operation is described as a function. This implies that the inputs and outputs of a net may represent *anything*. The network itself does not specify whether the inputs are apples and pears, classic music or trash metal, vestibular or proprioceptive inputs. This lack of specification also applies to recurrent networks: whereas we, as observers of the net’s output, are likely to recognize Mao Zedong, the activity might as well represent something completely different. The network receives no separate input about *what kind* of information it is processing, or which sensory modality it is working on. Neither does the network generate this knowledge by itself. It is only in the eye of the beholder that the net’s information traffic acquires any content.

But could this problem of input–output identification not be solved easily by taking the nervous system’s sensory receptors as our point of departure? After all, the peripheral nervous system has dozens of physical transducer types at its disposal to provide very specific inputs to precise anatomical target regions in the brain. Churchland and Sejnowski (1992) frame this view as follows:

How do neurons represent—anything? Because of their proximity to the periphery, and in view of the opportunity to correlate neuronal responses with controlled stimuli, the sensory systems are a more fruitful starting point for addressing this question than, say, the more centrally located structures such as the cerebellum or the hippocampus or prefrontal cortex. [...] Constrained by transducer output, the brain builds a model of the world it inhabits. That is, brains are world-modelers, and the verifying threads—the minute feed-in points for the brain’s voracious intake of world-information—are the neuronal transducers in the various sensory systems. (pp. 142–143)

So what could be the nagging issue that does not fit in? To address this question, we must go back about 350 years in history and revisit the labeled-lines hypothesis, launched about 200 years ago.

4.5 Labeled Lines and the Identification of Sensory Modalities

The labeled-lines hypothesis asks how inputs traveling from the sensory periphery to the brain come to be determined as coding particular sensory qualities, such as vision, touch, and smell. Historically, this problem can be traced back to John Locke's *An Essay Concerning Human Understanding* (1667). On the one hand, Locke asked how impressions and concepts arise in our minds and posited that humans are born as a *tabula rasa*, a white paper, devoid of innate ideas or built-in preconceptions about the world. All contents of our consciousness must therefore come from experience, which is ultimately or directly derived from sensory inputs. On the other hand, he argued that what we directly perceive are not objects (substances) themselves but rather their qualities as these are communicated via the senses. Two fundamental kinds of quality are distinguished: primary qualities comprise the solidity, number, movement, size, and shape of perceived objects, whereas secondary qualities refer to their color, taste, smell, temperature, and sounds. This distinction rests on the reasoning that primary qualities are inalienable from the objects to which they belong. According to Locke, we have no reason to assume that the perceived numerosity of objects in a group would be different from their actual number in the world. However, our perception of a hot bath steaming with lavender and eucalyptus aromas may differ from person to person, and there is no immediate or natural correspondence between these subjective, qualitative sensations and the physical properties of objects that cause them. Nowadays "qualitative sensations" are also referred to as "qualia" (see chapter 3), and here these denote all distinct (conscious) sensations in the various modalities and submodalities such as smell, taste, depth, touch, color vision, dizziness, nausea, or feeling tense.

At the time, Locke did not convincingly explain *why* or *how* the distinction between primary and secondary qualities would come about when trying to trace back the nature of sensory impressions to physical transduction principles. Is the distinction as fundamental as it seems to be? There are two principal reasons for denying this, the first of which relates to our knowledge of sensory transduction processes that was lacking in the seventeenth century. Smell, we now know, relates to the chemical structure of odorants that determines their specific binding to about 1,000–1,300 different kinds of genetically encoded receptors in the nasal epithelium (Buck & Axel, 1991; Mombaerts, 2004). We do not know how it is that the spike patterns traveling from receptors to the brain give rise to a sensation of smell, but at least we can say that the specificity of a detected odor corresponds to the specificity of the odorant's chemical structure and its binding to a multitude of olfactory receptors. The degree of "correspondence" between qualitative sensation and physical properties is

gradual and dependent on our empirical knowledge of the sensory transduction system. Similar accounts hold for other modalities.

The second principal reason is that one cannot maintain that some kinds of sensation may differ from person to person (and would thus be “subjective”) whereas others would not. Also primary “qualities” such as movement, size, and shape can be subject to illusion, likely with different strengths between human observers. And neurology has demonstrated that the subjective perception of motion or shape can go missing when brain areas are selectively damaged (see chapter 3).

The difficulty in making distinctions between “quantitative” and “qualitative” sensations is well illustrated by *depth* perception (cf. Boring, 1950). Depth can be quantified by making measurements with a yardstick or by counting your steps when walking through a room. Yet, depth perception can be said to add a real quality, or dimension, to two-dimensional vision, as its absence makes us see the world as flatter. As a submodality of vision, depth perception is neurally grounded in visual cortex neurons responsive to stereodisparity.³ Malfunctioning of this system does not completely deprive us of depth sensations because our visual field offers some remaining indicators of depth, such as perspectival lines toward a vanishing point on the horizon. Nevertheless, there is a qualitative difference between the sight of a perspectively correct painting and the real-life, three-dimensional visual experience of the scene the painting refers to.

In sum, quantifiability of sensations does not contradict their qualitative nature. The feeling or sensation of depth is as much of a “quale” (if we are forced to use this ugly term) as the smell of lavender. We can outline reasons why some sensations have a stronger “qualitative feel” to them than others (see chapter 9), but the point is that the problem of conscious representation is not limited to sensory qualities corresponding to the classical “qualia” or to Locke’s secondary qualities. It comprises the full realm of perception—including motion, spatial extension, and numerosity. By consequence, our perspective on the central problem shifts away from “qualia” to the more general problem of how *sensory modalities* are identified or coded by the nervous system.

Following Locke, it would take more than a century before more concrete, empirical hypotheses started materializing. In 1802, Young proposed his trichromatic theory of color vision, positing that fine color discrimination is not matched by an infinite number of particles in the retina, each of which would “vibrate in perfect unison with every possible undulation” (i.e., wavelength of light). Instead, he proposed that color vision is mediated by a limited number of light-receptive particles—for instance, responsive to red, yellow, and blue as principal colors. The new idea was that every unique sensation or experience is not matched by a specific type of receptor on the body’s sensory surface but can be a *composite* of the activity of multiple receptors. Nowadays, we take for granted that vision across the entire color spectrum rests on three types of retinal cone receptor, tuned to short (bluish), medium (greenish), and long (reddish) wavelengths of light, but at the time Young’s idea was revolutionary. However, Young’s theory did not make clear *how* the

activity of different photoreceptors could be combined to code a unique color sensation, and implicitly assigned the location where color sensations would happen in the retina, not specifying a role for the brain.

Charles Bell (1869/1811) and Johannes Müller (1838) adopted a more general perspective on the coding of sensory modalities, taking Young's theory into account and extending it to the brain and afferent nerves. Bell and Müller examined whether the nature of a sensation is determined by a specific pattern of activity emitted by sensory transducers on the body surface or by the distinct properties of that receptor. The first option, referred to as "pattern coding," is exemplified by a hypothetical photoreceptor sensitive to both red and blue light. It would transmit wavelength information by different activity patterns (e.g., a burst of action potentials in response to red light vs. a slow, regular train to blue light). The second option, referred to as "labeled-line coding," is the hypothesis adopted by Bell and Müller, and it received widespread support from subsequent research, most notoriously by Hermann von Helmholtz (1863, 1866) on the visual and auditory system. It postulates that it is the nature of the distinct sensory receptors, and the nervous pathways that relay their activation patterns to the brain, that determines the modality a sensation belongs to.

In his corresponding doctrine of "specific nerve energies," Müller (1838) formulated ten laws, of which two particularly relevant ones can be restated as follows: (3) "the same external cause [...] gives rise to different sensations in each sense, according to the special endowments of its nerve" and (5) "we are directly aware, not of objects, but of activity of our nerves themselves."⁴ Müller's "law of specific nerve energies," laid down in these ten postulates, seems now trivial to us, as we are all too familiar with a smack on the eye causing not only a tactile sensation but also a visual one. This law is also referred to as specific irritability (Bell–Müller) or adequate stimulation of nerves (Sherrington, 1906), as they have a unique relationship with the physical events by which they are excited.

Yet, something strange is observed as we steer our attention away from sensory periphery to brain. How could it be that we are aware of activity of our nerves *themselves*? Would this mean that we are conscious of the very spike trains running along the axons of our nerves? If so, why don't we simply perceive or feel action potentials instead of a visual image or stroke on our skin? Presumably we should take Müller's law to mean that somehow the brain converts or "translates" spike trains conveyed by nerves into percepts. But how could it do so in the absence of any *intrinsic* knowledge of where those spike trains are coming from? If all types of nerve fiber operate in the same way, conducting spikes to the brain, and the brain is not informed about the nature of the receptors from which spikes originate, the specific characteristics of each modality must arise from localized centers and operations in the brain (cf. Boring, 1950). Thus, the problem of modality identification has been relegated from the peripheral to the central nervous system.

Imagine a neocortical neuron, part of a sensory system of unknown modality. The neuron is quietly soaking in its bath of extracellular fluid, firing at leisure and receiving a few background inputs from neighboring neurons. Suddenly, a surge of electrical

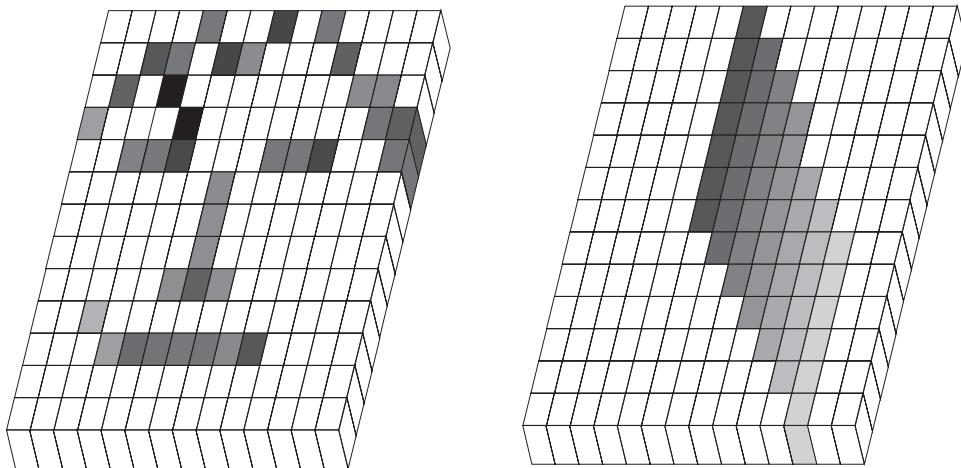


Figure 4.9

The brain's equivalent of René Magritte's painting "Ceci n'est pas un pipe." The key idea is that a group of neurons, each of which is represented by a voxel, may seem to generate, in its activity pattern (shown in gray tones), an unequivocal representation of a face (left), whereas by itself this activity pattern is not sufficient to code the visual nature of the information. Without further knowledge, the pattern may as well be generated in the auditory or somatosensory cortex. The activity pattern in the panel on the right illustrates that dynamic activity patterns are also insufficient in revealing the sensory modality of any input. In this panel the dark gray array of voxels represents the most recently activated neurons, with lighter gray tones for neurons that were activated earlier and earlier in the recent past. Such a time-dependent activation pattern likewise fails to reveal the sensory modality it belongs to.

activity arrives via the neuron's afferent axons, unleashing glutamate and GABA release and causing it to fire a burst of action potentials. What is this flurry of activity about? All afferent nerves work in the same way, exciting lower-order sensory relays which propagate the information to higher centers, all using basically the same types of transmitter and postsynaptic response. A single neuron—or a collection of neurons in one sensory center—does not receive information about the anatomical origins of the input it receives (see figure 4.9). Thus, the labeled-lines hypothesis fails to answer the question of why similar impulse patterns give rise to qualitatively, modally distinct percepts.

Likewise, the hypothesis leaves unsolved the problem how a neural network can identify or "recognize" the sensory modality fed into it.⁵ A classic net has no independent capacity or built-in knowledge to interpret the input it is processing. Churchland and Sejnowski's concept of brains as world modelers is an attractive one, but to propose that sensory receptors act as the "verifying threads" begs the question of how any brain area is able to verify, or identify, the kind of information it receives. If the brain has no means to label or recognize the brick stones, the substance for building a world model is lacking.

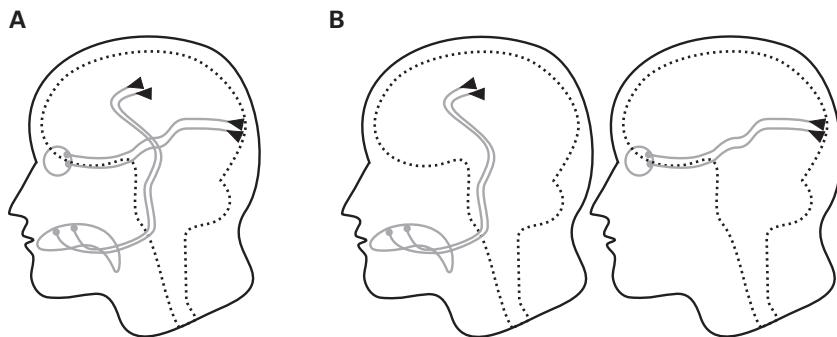


Figure 4.10

The “two separate brains” argument for the necessity of sensory integration in the brain. (A) The core elements of visual and gustatory signal propagation from retinal photoreceptors and taste buds to the main receptive neocortical areas in the human brain. Brain stem and thalamic relay stations have been left out for simplicity, and only two afferent lines, synapsing on postsynaptic cortical targets (black triangles), are shown per modality. If it is assumed that these basic pathways would be sufficient to endow the brain to see and taste, with a subjective ability to distinguish and experience sensations in these modalities at the same time, it should be logically possible to distribute the visual and gustatory pathways across two separate brains (B). Here, seeing and tasting should still be happening at the same time, but these sensations are now dissociated between persons—and yet the ability to experience these modalities distinctly and jointly should be retained. The upshot of the argument is that no sensory modality can be attributed to either system if such receptor–receiver pathways remain segregated and noninteractive.

Pursuing the labeled-lines hypothesis virtually to the level of absurdity, figure 4.10A illustrates the juxtaposition of visual and gustatory pathways we are comfortable with. Vision and taste peacefully coexist within our subjective experience, like they also cohabit neuroanatomically in our own brains. However, if their modality-specific wiring stays apart within a given brain anyway, why not separate the visual and taste modalities across two different brains (figure 4.10B)? What would it matter for the quality of sensations if the visual system was part of your brain, and I got the taste part? One possible answer could be this:

“Well, of course, if my brain were lacking any taste input, I would not be able to taste anything, and you would be blind!”

“Try to pinpoint the reason why a two-brain processing system would be fundamentally different from a one-brain system. Suppose that the two brains converged on the same trunk and had a common body. Would this unusual Siamese twin be able to taste and see? Why would the addition of a trunk matter if the relevant senses only depend on cranial nerves?”

To complicate the situation even further, imagine that the neurons in your visual cortex have been reaferented and now receive taste-bud input, without your knowing. Would you continue to see or switch to tasting? Even when two receptor–receiver systems are present

in the same brain but remain unconnected and noninteracting, there is no reason to attribute *any* sensory modality (or “quality,” experiential specificity) to them. As an area of sensory cortex appears not genetically predestined to become dedicated to a fixed modality (see chapter 2), the specificity of its processing should be based either on the statistical properties of its inputs or on the interactions between its processing modules with those for other modalities. Noting that an “input-statistics” view faces the *a priori* problem that numerical relationships and probability distributions do not disclose anything about the nature of information processed in a system, this matter will be further discussed in chapters 5 and 8.

For the moment we are left with the unanswered problem of modality identification. Perhaps a bit of consolation may be offered by the fact that signaling in distinct modalities is at least marked by the activation of *different sets* of neurons in the brain—neurons responsive to visual inputs in occipital cortex, sound-tuned neurons in the auditory temporal lobe, and so on. We can still maintain that differential activity patterns across the whole brain make up a *discriminative state* by which it can tell apart a state of receiving visual or auditory input. However, this observation does not specify how or why qualitatively different experiences should arise. In chapter 3 we examined the functionalist claim that the qualitative “feel” of conscious experience would be made up of reactive dispositional states, including beliefs and motor inclinations triggered by sensory input. But the brain can be stripped of these capacities and still sustain qualitatively distinct sensations. We could also adopt another view on discriminative states and claim that their perception-relevant aspects are purely made up by differential activity patterns across sensory brain areas, leaving structures for emotions, language, and declarative memory aside. However, this would yield such an unconstrained, general data representation model that it would lack sufficient power to explain how the richness of conscious experience, made up of so many distinct inputs distributed across modalities and submodalities, might arise. We could number all neurons from all sensory areas involved and compose a multimillion valued vector representing the activity per neuron:

[0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0,]

Next, we could go on with describing a different discriminative state, compute the inner product between this vector and the previous one, and establish how similar different states are. However, this numerical description does not bring us anywhere near the understanding of why brain activation patterns give rise to experiencing the smell of jasmine, the taste of a Brunello, or the sight of a Rothko. The problem with the concept of discriminative states is that it is too generic to explain what makes brain states so peculiar that they result in qualitatively differentiated experiences.

4.6 Why the Brain Is Also in a Chinese Room

Arguing against materialism, and more specifically against the claim that the brain and mind work like digital computers and computer programs, John Searle (1980, 1992) presented his famous analogy between a computer and a Chinese room. To recall the situation, we suppose that a native English speaker, unable to speak or understand Chinese, is placed inside a room and is given a set of instructions that tell him how to process inputs made up by strings of Chinese characters and convert them into outputs, also in Chinese. He receives the input symbols through a slot and passes the results of his conversion to the outside via the same slot. This configuration mimics a computer program trying to pass a Turing test by reacting to linguistic messages in fluent, intelligible Chinese. Searle asks, does the person inside the room (or anything else in it) understand Chinese, or is he conscious of the contents of the messages being passed in and out? Searle argues that neither the man, nor the instruction cards, nor the room as a whole is conscious and concludes that we cannot attribute a mental state to the room or to any similar symbol-manipulating system such as a computer. Therefore the brain must differ from a computer in that it possesses biological properties that are causal to having a conscious mind (cf. Searle, 2000).

The question of what makes brains, but not laptops, generate minds becomes even more acute if we contemplate figure 4.11, illustrating an important consequence of the modality identification problem⁶: *the brain also lives in a kind of Chinese room*. Via its spinal cord and cranial nerves, the brain receives messages, in the form of spike trains that come unlabeled and have no intrinsic meaning to the neurons that process them. Hence the name of this analogy: the cuneiform room. Central sensory neurons are analogous to an Englishman ignorant of Chinese, except that they do not even understand English. The “instructions” they operate on to convert inputs to outputs are described, at least primitively, by the activation and learning rules of equations 4.1–4.3. As much as neurons do not have a key to decipher the sensory inputs they receive, they do not understand the effects their outputs will produce throughout the body, interacting with the environment. We cannot attribute sensory content to them a priori and without further assumptions. And yet—we do know from our own experience that our mind–brain *does* pull off the trick! We do have extremely diverse, rich experiences and savor the world in all of its enjoyable or sometimes disgusting qualities. How can this be if our brains are locked inside a box that only lets through nerve fibers with electrochemical pulses of activity?

Here I argue that a renewed search for physicochemical properties, as a potential basis for explaining mind–brain properties, would not be fruitful. Even if such properties were to be discovered, they would be unable to explain how consciousness is generated. The building blocks of the brain are not fundamentally different from those composing inanimate systems in nature. A pursuit of novel atomic or molecular properties would bear the brunt in much the same way as vitalism was doomed to decline in the early twentieth century. Whereas vitalism clung to the idea that living systems possess some extra

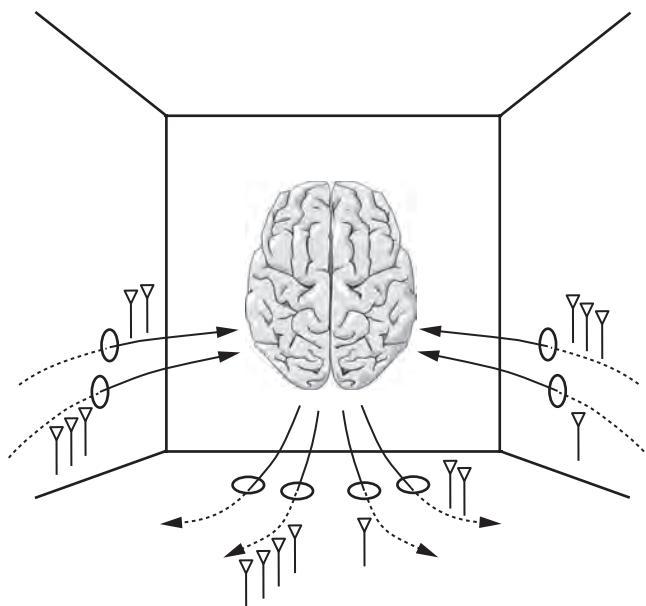


Figure 4.11

The cuneiform room: an analogy in which the brain, with its peripheral inputs and outputs, is situated like the operator in Searle's Chinese room. The cuneiform symbol represents an action potential. The spinal cord and cranial nerves have not been separately drawn for simplicity.

properties beyond and above what was known about biomolecules from physics and chemistry, empirical biology strode along to unravel the principles of how enzymes, DNA, and RNA work, making vitalism obsolete. This does not imply that current physics, chemistry, or neurobiology would be sufficient to explain consciousness, nor would I defend the idea that we should resort to eliminative materialism or functionalism. In the next chapters we will probe the organizational principles of the brain, that is, the principles that may guide information processing and sensory integration at other aggregate levels than single neurons or circuits. Despite their limitations, neural networks may offer inspiration in these explorations.

5

Networks and the Problem of Panpsychism

Look at a stone and imagine it having sensations. One says to oneself: how could one so much as get the idea of ascribing a *sensation* to a *thing*? One might as well ascribe it to a number!

—Ludwig Wittgenstein (1953)

5.1 Inanimate Networks and the Problem of Panpsychism

In chapter 4 we witnessed emergent properties of neural networks enabling basic cognitive operations. One may criticize these basic models for many reasons—for instance, their inability to sustain a continuous temporal “stream” of representations analogous to William James’s (1890) stream of consciousness. However, some of these deficiencies can be fixed; the lack of temporal continuity is addressed by more recent models deploying continuous attractors. These networks do not converge to a “fixed” final state but evolve from one state to the next (Tsodyks, 1999; Wu & Amari, 2005; McNaughton et al., 2006). At this time, however, a crucial remaining deficiency of network models is their inability to modally identify or recognize the inputs they receive, or their internal representations. It remains a mystery as yet how our own brains can decipher the cuneiform messages they receive from the periphery into a rich palette of modally specified experiences.

Here I will present a different perspective on the same problem by asking to what extent current models resemble nonliving systems. If we find our brain models trustworthy at first glance, but recognize strong analogies with other complex systems found elsewhere in nature, we face the question of whether our models are still too limited or simply inaccurate. If not, we may have to accept the position of panpsychism.¹ The alternative option holds that current models may be underconstrained: they may be too generic and should be specified further to find out whether they can genuinely generate cognitive phenomena we cannot ascribe to inanimate systems throughout nature. In addition to examining further constraints, we may go forward by questioning the conceptual approach underlying the models. Let me first illustrate the approach by comparing network models with a system that appears utterly nonconscious at first glance.

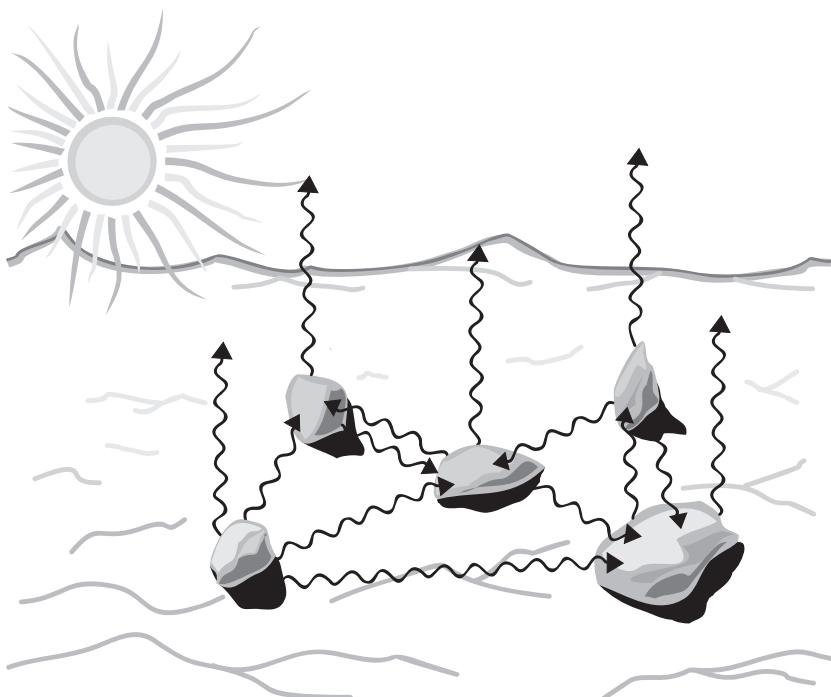


Figure 5.1

A group of rocks displaying neural-network-like interactions. When the rocks are hit by sunlight, they heat up and emit infrared radiation (arrows) to rocks in their vicinity. Not all interactions are shown. Impinging sunlight and upward emissions symbolize sensory input and network output, respectively. Rock temperature is analogous to neural activation state and changes in rock-to-rock infrared light transmission resemble adaptation of connection weights.

According to functionalism and some other positions, mind–brain functions are realizable in multiple ways; they can be implemented in different physical media. If we find neural networks credible enough as a starting point, we should ask whether they could be implemented not only by biological neurons made of the greasy, fluffy matter inside our skull but also by distinctly nonliving systems of interacting elements. Thinking of a group of rocks bathing in sunlight (see figure 5.1), it might seem ridiculous at first sight to think of them as implementing a cognitive operation, but I will argue to the contrary, at least for now.

Each rock is struck by sunlight and thus warms up at sunrise and cools down at sunset. Sunlight plays the role of *sensory input* to this rocky network. As the boulders heat up, they gradually emit more infrared light that radiates into the environment, including to nearby rocks. A rock's surface temperature is analogous to a neuron's graded activation state, and its emission of infrared light models a neuron's graded output, reaching other elements

in the network. A rock may be oriented and placed relative to another rock such that it receives a lot of the “sender” rock’s infrared signals. The strength of rock-to-rock transmission varies, much like the synaptic weights in a neural network. The collective output of all rocks is the sum of infrared signals issued by all rocks into the wider environment.

What kind of cognitive operation would this network be able to perform, if any? We could dryly measure the intensity of all infrared signals emitted by the rocks into their surroundings, and relate this to the intensity of impinging sunlight. This would elevate us to the remarkable task of classifying the sunlight input into categories such as “cool,” “luke-warm,” and “hot.” A human observer would have no problem “reading out” the network’s output and “decoding” from it the intensity of sensory input. Does this silly exercise expose a deep flaw in the usual business of network modeling, or is there something missing in the analogy?

Our approach may be too simple; the extraction of only one input parameter may be too rudimentary a task for what mind–brain models should be capable of. This argument is countered by taking more elaborate measurements. Given sufficient rock-output readings taken from many locations on a hemisphere overarching our rock formation, it should not be difficult to decode the *direction* of impinging sunlight next to its intensity. Sunrise will on average elicit stronger infrared emissions from stones facing east than those facing west.

It can be argued that the rock formation is too passive compared to a living, spiking neural network. Until now the rocks lack a recurrent architecture allowing outputs to be reprocessed (see figure 4.7), which would otherwise result in a dynamic evolution of network states for sequential processing and recognizing complex inputs such as images (Churchland, 1995). However, we can easily picture the boulders as grouped together inside a cave, some of them lying close to the entrance and some further down, into the darkness. If sunlight is reflected downward by the rocks close to the entrance, this light will hit a deeply lying rock facing that very entrance again, resulting in “feedback” connections from that rock to the “input units” close to the cave’s orifice. We are in fact dealing here with an ultrafast, optical network. The objection of the rocky network’s being passive cannot be upheld on account of a lack of reverberatory dynamics.

Is the analogy flawed because our rock formation lacks the capacity to learn and memorize? If “memory” is the capacity to retain previously acquired information, the rocks can be claimed to have memory-like properties: they can have a persistent activation state (rock temperature) as a cloud drifts before the sun. Sunset is accompanied by a slowly decaying “memory” of light accumulated during daytime, stored as thermal kinetic energy in the rock’s crystalline structure. But this type of memory *seems* passive—or can it be claimed to depend on activity-dependent modification of connection weights? Hebb’s rule, for instance, prescribes that connections are strengthened when both pre- and postsynaptic elements are consistently coactive.

Unsupervised models can be trained based on the statistical structure of their inputs. When two inputs consistently correlate with each other, this may lead to strengthening of

the connection between the elements responsive to each of them, as in a Hopfield network. In our rocky network, a change in sunlight input statistics may similarly affect the way boulders “communicate” with one another. If the sunlight is veiled by a deck of clouds, the input statistics will be dominated more strongly by diffraction than when it is emitted from a point-like source. Patterns of communication will become different, not only because of altered activation state but also because heating becomes more spread out across a rock’s outer surface, creating more sender-and-receiver connections, resulting in a more diffuse communication with some persistence over time.

I readily admit this configuration of rocks will not truly “learn” in the same way animals and humans learn, but the point is to show how hard it is to actually prove any formal distinction between a network model and a system many people would consider inanimate and utterly noncognitive. To illustrate the point about learning more poignantly, think of a stretch of lovely, warm beach sand that has just been washed over by a sea going into low tide. You walk across with bare feet, leaving behind deep footprints in the sand. As the interstitial spaces between the sand grains contain sufficient water, they retain the shape of your sole and toes. Why not claim that the stretch of beach forms a network of sand particles, which collectively learn and store a memory of your foot? The pattern of your foot’s underside is “imprinted” by changing the configuration of sand particles, retaining its structure by cohesive forces that are their connection weights. By the brute force of gravitation, plunging your foot in the sand reconfigures the connection weights, preserving your footprint for hours. True, this kind of network has no capacity to complete patterns or to store multiple patterns at the same time, but in principle the cohesive forces could be remodeled so as to implement the plasticity of a Hopfield net.² If your foot would be only partially imprinted in the sand, this plasticity would enable the beach to pattern-complete your entire sole. Surely it would be a magical experience jogging along a beach that reproduces your entire, personal footprint—but would this convince us that the beach has mind-like properties?

5.2 Neural Network Models: In Defense of Panpsychism?

Many people would find it curious—or even repugnant—to think of rocks and beaches as being conscious. However, regardless of intuitions, the question to ask is this: on which rational grounds is this judgment made? The rocks display no behavior indicating they would understand the photic inputs they are processing or would be able to act on them. They do not express any verbal behavior or body language suggesting their wakeful, active processing of inputs. Nonetheless, in chapter 3 we saw how the lack of overt behavior in locked-in patients is inconclusive for judging their state of consciousness. When presenting neural nets as models of mind–brain operations, we set no *a priori* demands that should be implemented in a behaving device such as a mobile robot. Network modelers have good reason to facilitate the emulation of cognitive operations by omitting such demands, which

would slow down progress in the field by sensor and actuator problems. This implies we should not consider the lack of overt behavior in our rock and sand networks a crucial omission. On its own, lack of overt behavior does not provide decisive ammunition against panpsychism.

A more valid objection follows from the difficulty in ascribing *representations* to rocky networks, as we infer these from our own experience, psychology, and neuroscience. It appears meaningless to think of what it would feel like to be a bunch of stones bathing in sunlight. We strongly feel that they cannot “know” anything, do not “represent” anything except for the physical shapes and configurations they assume passively, and have no “inner life”—no internal, self-organizing dynamics, capable of dreaming or imagination. The sand grains retaining your footprint are only holding a “memory” in that their physical configuration directly forms a mold of it—a mere re-presentation in the sense of an oil painting. This contrasts to the notion of consciously experienced memory as a re-creation or reconstruction of the original event, relying on active retrieval of information stored in “dormant” form, and on re-integration of scene elements and imagery (cf. Neisser, 1967; Marcel, 1983; Kosslyn et al., 2001; Daselaar, Porat, et al., 2010; Huijbers et al., 2011).

Powerful as this intuition may be, it yet fails to clarify why a neural net may possess cognitive properties whereas a network of rocks would not. The sluggish dynamics of our rocky network do not prove that the network has no mind-like properties whatsoever: it may just have a very “slow” mind, scaled to phenomena such as the 24-hour rotation of the earth around its axis. The difficulty in pointing out a crucial difference between neural and rocky nets brings us to a dichotomy: if we assume that all phenomena we encounter in nature are physical by nature, should we then embrace panpsychism, as argued by some contemporary philosophers and neuroscientists (Seager, 1995; Strawson, 2006; Koch, 2012)? Or should we qualify current neural network models as insufficient or plainly wrong when comparing them to real brains?

Here I will argue against the sufficiency of current models. In chapters 6 and 7 we will survey brain systems that can be reasonably linked to, or dissociated from, consciousness. A system not linked to consciousness (which bears, at best, an indirect relationship to it) will be observed to consist of the same neuronal elements and same types of connections and neurotransmitters as found in systems that can be associated with consciousness. Both classes of brain system can be defined as “neural networks” in a broad sense, but one class appears indispensable for consciousness, whereas the other class can be missed (although its loss gives rise to other problems, such as of motor control). This logic prescribes that generic network models fail to incorporate some essential properties that distinguish brain systems for conscious *versus* unconscious representations.

In his review on the biological basis of consciousness, Searle (2000) asked what the causal powers of the brain are—what distinguishes them essentially from other complex systems in nature that do not give rise to consciousness? This question is well-taken, but we do have to bear in mind that these “causal powers” may not be identified by studying

lower physical–chemical properties such as the ion channel composition of neuronal membranes but may relate to higher-level functional principles of systems organization. If our search for “causal powers” zooms in exclusively on cellular or subcellular phenomena, we likely end up in the same deadlock position in which vitalists such as Hans Driesch found themselves in the early twentieth century when holding onto undiscovered, mysterious powers that would distinguish living cells from inanimate structures, beyond their normal biochemistry (Mayr, 1988). Thus, our search for causal powers should be guided by differences between systems *within* the brain, studying which *systems properties* endow only a subset of brain structures with power for conscious representation.

However, even the brain systems linked to consciousness can assume nonconscious states—such as most episodes of deep non-REM sleep³ or anesthesia. Their anatomical structure remains like it was before a state transition occurred, so it must be a particular *mode of activity* that is important for generating conscious representations. Together, the combination of anatomical, physiological, and psychological observations raises strong arguments against panpsychism. Even if the adjunctive “pan” is confined to the nervous system, panpsychism does not hold.

Abscribing a mind to almost any multiparticle system in nature would, in the end, lead to such a generic and nonspecific use of the term that it would lose its meaning. If just about anything in nature qualifies as being conscious, why not abolish the whole concept? “This rock surely seems conscious but is actually sleeping” is a nonsensical statement. We find fellow humans and certain animals around us that display sleep–wake cycles and other behaviors we may reasonably associate with consciousness. Moreover, their brains show great anatomical and physiological similarities to our own brains. But this similarity does not hold for all organisms or complex systems in nature, let alone human-made devices such as mobile phones.

5.3 Computational Functionalism and Semantic Networks

When tested against panpsychism, do other models of mental processing fare any better than classic neural networks do? In chapter 3 we encountered a variant of functionalism advocated by Dennett in discussing qualia. More broadly, we saw how classic functionalism posits mental processes as entities fulfilling causal functions intermediate between inputs and actions of an agent (Putnam, 1975a; Armstrong, 1980, 1993; Lewis, 1980a,b). As a leading metaphor, a digital computer processes incoming numbers, images, or symbols by subjecting them to a (mathematical or linguistic) transformation (Newell, 1990). This process is realized by the application of certain operators according to fixed computational rules. A mathematical program, for instance, implements an algorithm that may sieve out prime numbers. In the linguistic domain, the rules of engagement concern a grammar’s *syntax* as applied to words, chunks of words, or other entities carrying meaning or semantic content (Chomsky, 1957).

As founder of a more cognitive approach, labeled “psychofunctionalism,” or “representational theory of mind,” Fodor (1975; see also Pylyshyn, 1980) denoted these linguistic entities as representations. They refer to real-life contents such as a specific person or object, a category of objects, actions, properties, and more abstract concepts. Fodor argued that mental processing should be envisaged as the processing of messages that are passed on from one functional module in the brain to another, and hence this view is also called “computational functionalism.” The language for coding these messages would be a “*mentalese*,” or “language of thought,” a system of symbols and internal-grammar rules for computing inferences from premises and other syntactical operations. How this “language of thought” exactly differs from a common language such as Spanish or from computer-programming languages has not become clear, but here we will merely probe whether these symbol-processing *semantic networks*, envisaged by computational functionalism, are susceptible to panpsychism.

Similar to neural networks, a semantic network consists of *nodes* and *connections* between the nodes, but now each node represents a meaningful item (a noun, verb, or adjective in the example of figure 5.2) whereas the connections symbolize specific relationships between two items (e.g., properties of a noun; categorical relationships; types of action; relationships of possession or membership; logical implications). Importantly, a “node” in a neural net is not fixed on a particular meaning or content (e.g., “chair”). In

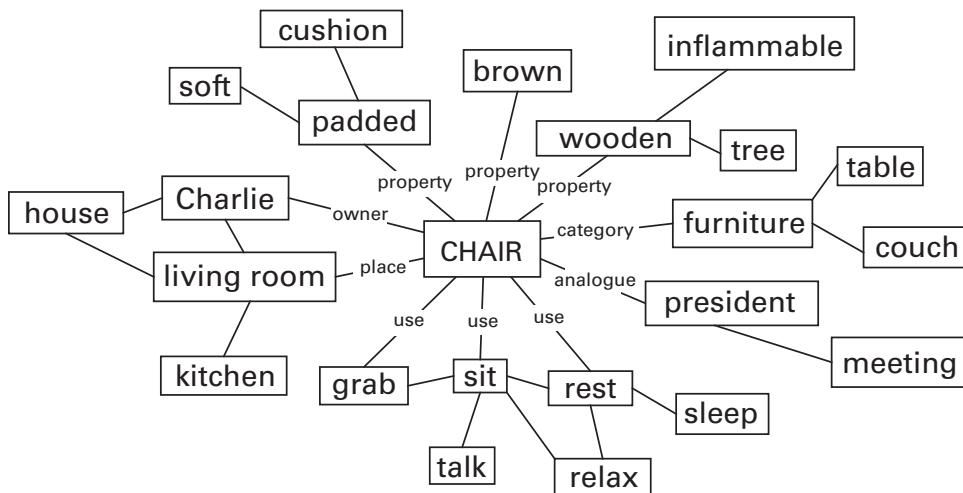


Figure 5.2

Example of a semantic network centered on the noun “chair.” Note that different types of connection define the relationships between “chair” and nouns, verbs, adjectives, or personal names. Types of relationship have only been denominated here for items having a direct connection to the central node. Loosely based on Collins and Quillian (1969).

contrast, nodes in a semantic network are content fixed and interact by heterogeneous types of connection indicating different kinds of relationship. Finally, a unit in a neural network may represent a single neuron or group of functionally similar neurons, whereas a semantic item is thought to be coded by a higher aggregate unit, not by a single neuron.

According to Block (1980b), neural networks—or functionalist constructs in general—can be compared to a “Chinese nation.” He supposed that every Chinese citizen could carry out the same function as a node in a network would do. Every citizen accepts messages from fellow citizens, carries out some processing step, and transmits its output to others. If the brain works as a functionalist network of comparable size and complexity, it follows that the nation of China must be conscious. This is not a reasonable thing to conclude, Block argues; hence functionalism fails as a model of the mind–brain. One of the implications is that, if functionalism would apply, many other systems besides the Chinese nation would qualify as entities generating consciousness, leading to panpsychism.

Block’s argument has been criticized in that it is not straightforward to formally deny some type of consciousness to the larger aggregate of the Chinese population. Perhaps some type of conscious representation *is* possible at a supracitizen level given a proper arrangement and activity mode of all citizens (cf. Lycan, 1987; Van Gulick, 1989). I find this counterargument hard to refute: if the behavior of all neurons in the brain is mimicked *exactly* by all citizens of China or of our entire planet (receiving innocuous excitatory and inhibitory potentials across their bodies and transmitting real-time action potentials along electric cables), why would this precise copy of the brain *not* give rise to consciousness? Here the argument of multiple realizability cuts both ways.

Returning to semantic networks, it is easy to plug the name of an item or property into a node, but if the model is to emulate a mind–brain entity, it is unclear how the meaning of any item would arise. Network nodes are assumed to process messages and transmit them to each other, but how these messages come to refer to percepts and concepts about the outside world or our own bodies remains obscure. Assigning meaning to different types of connection in a semantic network suffers from the same problem: intentionality does not arise for free. Without a mechanism to associate “pink elephant” with an experience of the color pink and an image of an elephant, it remains enigmatic why the symbol set “pink elephant” could not be replaced simply by an “x” to denote an arbitrary variable in a mathematical function.

Implicit to psychofunctionalism and its version of semantics is the notion that a symbol’s meaning is coded by the *network* of connections it has with other symbols. “Pink elephant,” for example, means nothing else than the composite of connotations the term has—namely, with a color, an animal species, a restaurant bearing this name, and so forth (see also Crick & Koch’s [2003] “penumbra” hypothesis for a related idea from neuroscience). However, because each of the associated items derives its meaning from connotated groups or symbols, an infinite regression of associations arises, with meaninglessness spreading out across the network like an ink blot—unless, perhaps, further constraints are

applied. Another strategy has been to assume that symbol meaning is derived from properly hooking up the formal symbol system to the external world (Fodor, 1980, 1985). At least part of a semantic network would be connected to the sensory periphery via a “lifeline” providing real-world meaning. However, rather than solving the problem of conscious representation, this approach comes down to a restatement of it: how *does* the brain convert spike trains arriving from the periphery into consciously experienced, meaningful content?

Harnad (1990) labeled this question the “symbol grounding problem”: “how can the meaning of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?” (p. 335). His solution is a hybrid scheme combining features from neural connectionism and symbolic architectures (see Pomerleau et al., 1991, and Sun & Bookman, 1994, for related approaches). Symbolic representations, used in a system of formal symbols, would be grounded in nonsymbolic representations of two types. The first type Harnad labels “iconic representations,” which result from the sensory coding of concrete objects and events and are comparable to one-shot, real-life episodes. The second type are categorical representations, formed by neural structures classifying singular experiences into larger classes by extracting invariant features. Connectionist components of his model serve to extract categories from specific, iconic examples.

Here, the meaning of symbolic representations hinges on categorical representations such as that of a horse, and the most elementary symbol in this case is simply the name of a category (“horse”). Neither symbolic nor categorical representations, however, can be interpreted to have meaning on their own. Meaning would only arise if the names of categories, grounded via iconic representations, were combined into propositions about category relations. For instance, a more complex symbol arises when combining elementary symbols, such as when a “horse” symbol is combined with “stripes”: the proposition is that a “horse” with “stripes” is a “zebra.” The symbol is grounded here because iconic, sensory “examples” of horses are transformed—by way of a trained connectionist network—into a categorical representation. In this hybrid system, symbol meanings would be “not just parasitic on the meanings in the head of the interpreter, but intrinsic to the dedicated symbol system itself” (p.345). However, it still remains mysterious how symbol meaning comes about.

Harnad admits that “iconic representations no more “mean” the objects of which they are the projections than the image in a camera does,” bringing us back to the observer who must attribute meaning to the image, rather than a symbol-processing system’s doing so. As it remains unsolved how iconic representations acquire meaning, a firm basis for attributing intentionality to categorical and symbolic representations appears to be lacking. Already at the basic level of sensory analysis that underlies iconic representations, the scheme faces the modality identification problem.

In conclusion, various computational functionalist or “symbolic” approaches have not offered a convincing solution to the grounding problem (for further discussion, see

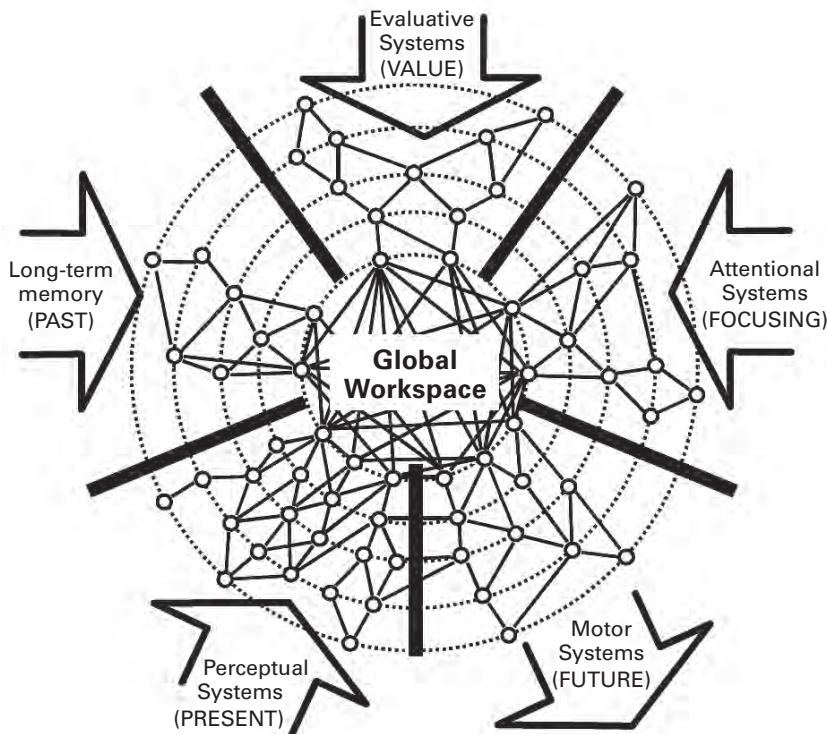
Smolensky, 1988; Barsalou, 1999; Marcus, 2001), leaving panpsychism standing as a problem as much as it was for neural networks.

5.4 Global Workspace Theories and Access Consciousness

Work from cognitive psychology and neuroscience has converged on theories of consciousness that differ from both classical neural networks and the “psychofunctionalist” approach of Fodor and Pylyshyn. Departing from Mesulam’s (1998) work, DeHaene, Changeux, and colleagues adopted a neural approach to a “global workspace theory” (see figure 5.3; DeHaene et al., 1998; DeHaene & Naccache, 2001; DeHaene & Changeux, 2011). Here, primary and secondary sensory areas feed incoming information into a central neural domain where it can be efficiently broadcast to widespread functional modules elsewhere in the brain, mediating general cognitive and executive functions (such as planning, attentional selection, working memory, and emotional evaluation) and other sensory modalities. The “workspace” acts as a hub or informational “marketplace” and is at the center of concentrically arranged circles that incorporate modality-specific processing at the periphery and more integrated processing toward the core. The central areas, including prefrontal, cingulate, and parietal cortices, do not necessarily lie close together in brain space because information can be rapidly trafficked across large distances via corticocortical fibers. Once an input has passed the lower stages of sensory and temporal cortex, which are considered to perform “preconscious” processing at this point (DeHaene et al., 2006), a salient input will penetrate the higher associational areas. Here, the signal may pass an “ignition threshold,” thereby unleashing a widespread broadcasting of the signal into brain structures in the inner ring and beyond.

Inspired by cognitive psychology and classical AI (Newell, 1990), Baars already proposed a somewhat comparable global workspace theory in the early 1980s (Baars, 1983, 2002). To solve the problem of how nodes in a widespread network communicate efficiently and get access to relevant information, he also envisioned a central stage receiving and distributing information from and to more specialized modules. Implicit to this proposal is the claim that the global workspace carries the contents of consciousness, whereas modules outside it process information unconsciously. The core idea of Baars’s theory—that consciousness serves to “facilitate widespread access between otherwise independent brain functions”—is that consciousness does not serve primarily to *represent* information (in a form we characterize phenomenologically as qualitative) but to solve the *logistic* problem of information exchange and access.

Scrutinizing global workspace concepts, Ned Block distinguished two kinds of consciousness, namely, “phenomenal” and “access” consciousness (Block, 1990, 2005; see also Baars, 2002). These terms delineate the phenomenal content of our subjective experiences vis-à-vis all cognitive processes operating upon and using this content, such as for verbal judgment, perceptual interpretation, attention, working memory, or motor decisions.

**Figure 5.3**

Global workspace model proposed by DeHaene et al. (1998). Scheme is based on a diagram originally introduced by Mesulam (1998). Progressing from the periphery toward the center, the concentric circles illustrate how information from specialized functional modules may be integrated and accessed in the global workspace at the core. The direction of information flow can be mainly toward the center, as illustrated for long-term memory and evaluative, attentional and perceptual systems, or primarily outward, as for motor systems. Copyright (1998) National Academy of Sciences, U.S.A., reproduced with permission.

When cast in this dichotomy, global workspace theory mainly addresses access consciousness and has less to say about the neural basis of phenomenal awareness.

We need to scrutinize whether access consciousness truly represents a different kind of consciousness than phenomenal consciousness. Should we speak of two kinds of awareness, with contents we are conscious of passing from a phenomenal stage to a stage where the information is used for verbal reporting, facial expression, and so forth? For various reasons it appears more parsimonious to consider access consciousness not as a separate form of consciousness but as a conglomerate of executive functions operating on ongoing sensory representations: allocation of attention, storage in working memory, behavioral decision making, and so on. One of these reasons is that, in order to learn what a subject was aware of, one will be forced to probe a subject's mechanisms for verbal behavioral reporting (cf. Cohen & Dennett, 2011). In my view the field will benefit more from a

distinction between neural systems for mediating different aspects of consciousness as these are studied in different ways (e.g., sensory-representational aspects, behavioral manifestations, interaction with memory and attentional systems) than from a strict phenomenal–access dichotomy which is hard or impossible to demonstrate empirically.

The network models of DeHaene and colleagues underpinning global workspace theory are capable of solving advanced cognitive tasks such as the Tower of London planning task but are nevertheless genuine, classic neural networks trained by reinforcement learning (DeHaene et al., 1998). Despite their complexity, they are thus subject to the same problems maintaining the gap between classic networks and conscious systems, including the problem of panpsychism. More questions pop up as we think about the idea of a central workspace. For instance, Baars (1983, 2002) argued that integration of information emitted by various modules must occur in global workspace for a conscious experience to occur, but it does not become clear how this works to raise nonconscious information from the periphery to a conscious level (cf. Chalmers, 1995). Besides the critique that especially Baars's approach resembles a Cartesian theater (Dennett, 1991; Rose, 2006), where a viewer or homunculus must be invoked to “bring it all together,” the approach can be further questioned by asking whether a “central workspace” is necessary to ensure proper access to information—the original rationale for the theory. Modern communication architectures such as the Internet demonstrate that information can also be rapidly exchanged across distributed nodes or decentralized hubs in a “small-world” architecture (Watts & Strogatz, 1998).

So far, we have seen few reasons to feel encouraged that cognitivist or global workspace hypotheses solve the problem of panpsychism. This is not meant to state that the architectures are irrelevant to cognition, because they do capture a number of processes such as learning, distributed representation, and emergence.

5.5 Entropy, Complexity, and the Ubiquity of Information

Following earlier work in behavioral and cognitive psychology (Hebb, 1949; Neisser, 1967; Marcel, 1983), AI (Newell, 1990), PDP (Minsky & Papert, 1969; Hopfield, 1982; Ballard et al., 1983; McClelland et al., 1986; Grossberg, 2007), and systems neuroscience (Allman et al., 1985; Ts'o et al., 1986; Gray et al., 1989; Von der Malsburg, 1995; Felleman & Van Essen, 1991; Blasdel, 1992), Tononi, Edelman, and colleagues developed a “dynamic core hypothesis” of consciousness deriving its main features from information theory. Instead of focusing on which particular brain areas contribute to consciousness, Tononi and Edelman (1998) emphasize the search for neural *processes* characteristic of conscious experience. They consider a conscious state as a rich, complex, and unified experience coded by a large cluster of neuronal groups (the “dynamic core”) that stands out because of its high complexity and integration of information. This core typically includes regions in the posterior and anterior corticothalamic systems, involved in perceptual categorization and executive

functions, respectively. The core is “dynamic”: it is assembled and dissolves across the fast time scales characteristic of the genesis and fading out of conscious experiences. The system also permits a flexible membership of neuronal groups (e.g., a perceptual switch in viewing an ambiguous figure invokes participation of different groups).

The key feature of the hypothesis is that a conscious state is explained by the neural processes of integration and differentiation, which are both quantifiable properties. “Differentiation” means the availability of a vast repertoire of possible states a conscious system may assume. That any particular conscious experience occurs out of billions of alternatives means that it reflects an exceedingly high amount of information. According to Shannon (1949), information is defined as reduction in uncertainty, and the transition from an unspecified state to a specific experience enormously reduces the uncertainty about what is represented by the system. “Integration” refers to wholeness or unity of a perceived scene: we view it from a single perspective and do not experience independent components or different interpretations of it at any given time.

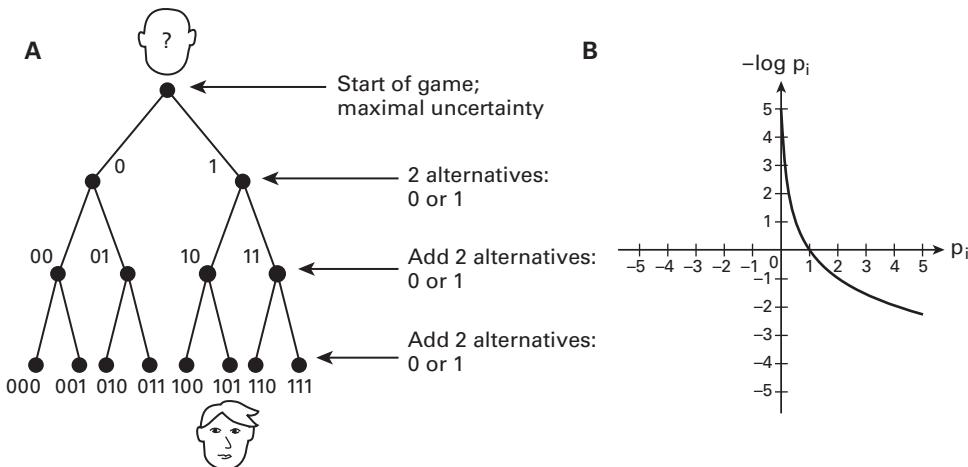
To determine whether a neural process is integrated, Tononi and Edelman used Shannon information to devise a measure of functional clustering. Let me first explain the scope of information theory, and how it relates information to reduction of uncertainty. Imagine you are playing a guessing game (similar to, e.g., Guess Who?) in which you keep a particular object (e.g., a toothbrush) in mind and another person tries to guess what it is by asking consecutive yes/no questions. At the start, uncertainty is maximal as no object information is available. After answering the question “Is the object living or nonliving?,” a first piece of information is resolved. When answers to questions like “Is the object used at home?” are given, the uncertainty about object identity decreases further, until the estimated probability of the object’s being a toothbrush is approaching 1 and a definite guess can be made. If the choice between two alternatives is equally likely (probability $p = 0.5$), we say that each step in the sequence conveys one bit of information (bit: binary digit, 0 or 1; see figure 5.4A).

Answers to three binary questions raise a total of eight possible alternatives or “states”: $8 = 2^3$. For binary choices this calculation is generalized to the following:

$$I = ^2\log K \quad (5.1)$$

with I being the amount of information (3 bits), and K the number of states (or alternatives) to be coded. The choice of 2 as base of the logarithm is derived from working with binary decisions. This choice is admittedly somewhat arbitrary.

Shannon information is closely related to the *entropy* in a multistate system. This concept originates from thermodynamics, where it is classically related to the amount of energy that cannot be used in work done to accomplish a process (e.g., dissipating heat in a car engine; as an informal shortcut, entropy resembles the amount of “disorder” in a system). Information theory defines the entropy H as the amount of information that you would miss if you wanted to know the value of a random variable of interest. In the example above, entropy

**Figure 5.4**

(A) Tree diagram illustrating the concept of information as reduction in uncertainty. At the top of a binary decision tree, an unknown face is presented, corresponding to a situation of maximal uncertainty. When a guessing game is played and yes/no answers on personal features become available, uncertainty decreases and one bit of information is added at each answer. After three answers, specifying eight (2^3) possible faces, enough information has been gathered to identify the target (i.e., the face coded by "101"). (B) Mathematical relationship between p_i , the probability of a world state i occurring, and $-\log p_i$, the quantity forming the core of equation 5.3 to compute information. Only the range [0, 1] applies to values p_i may assume.

can be computed as the log of the number of alternatives in the system: H equals 3 (and thus equals I), with a total of eight possible states (Rieke et al., 1997).

We can generalize equation 5.1 to a situation with states that may have different probabilities:

$$I = ^2\log K = -\sum_{i=1}^K \left(\frac{1}{K} \right) \log \left(\frac{1}{K} \right) \quad (5.2)$$

in which the base of the logarithm has been omitted for convenience. We have now taken the logarithm over the inverse of the number of states, divided the outcome by K , but also summed up across all individual states of the system, indexed by i . Realizing that the probability p of any state occurring (p_i) is inversely proportional to the total number of alternatives, we now obtain a more useful equation:

$$I = -\sum_{i=1}^K p_i \log p_i \quad (5.3)$$

When an event becomes increasingly unlikely, $-\log p_i$ increases much more steeply than p_i decreases (see figure 5.4B). This implies that very unlikely states hold more information than very common ones. For instance, when the police interrogate a witness at a crime

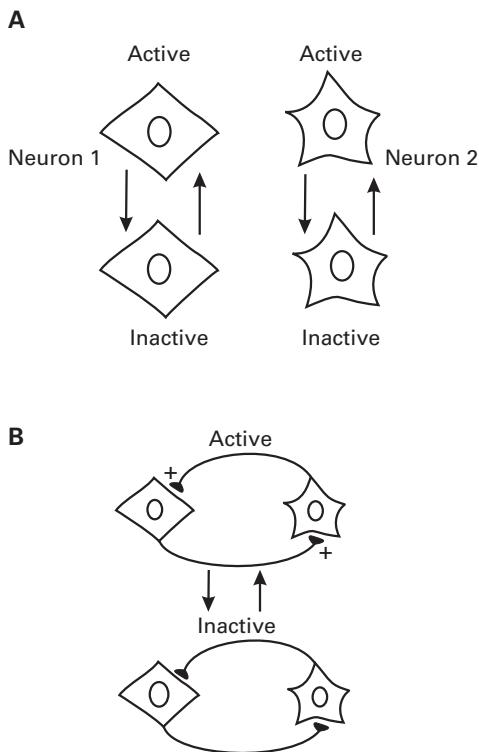


Figure 5.5

The concept of mutual information applied to neurons. (A) Two unconnected neurons independently assume active or inactive states; the total number of joint possible states is four (entropy: 2) whereas the number of states per neuron is two (entropy: 1). Mutual information is zero in this case. (B) If two neurons are connected such that activity of one neuron results in the other neuron's also being active, the total number of joint possible states drops to two (entropy: 1), rendering mutual information positive. Note that a transition from an active to inactive state can only be achieved if an inhibitory external input is assumed to occur (not drawn here).

scene, answers like “the robber had blond hair, wore glasses, and had a moustache” convey much more information than just saying that the person was a man.

The scope of information theory is limited to the statistical relationships elements in a system have with each other, such as when a particular event (dark clouds gathering) has a certain likelihood of being followed by another event (raining). This notion of statistical “connectedness” can be applied to real-world stimuli, brain events such as EEG signals and spike trains, symbols in a computer language, or words in a spoken sentence. Hence, Shannon information is useful for quantifying *syntax* and grouping of system elements but has little to say about semantic content (Shannon, 1948, 1949; Eggermont, 1998).

Returning to Tononi and Edelman’s hypothesis, the integration in a neural system is quantified here by computing mutual information (*MI*) as a measure of the statistical

dependence between its elements. If we have two complementary subsystems A and B which are part of the overall system AB , then

$$\text{MI}(A;B) = H(A) + H(B) - H(AB), \quad (5.4)$$

where $H(A)$ is the statistical entropy of subset A . Mutual information captures the entropy that is shared between the two subsystems A and B ; it represents the *surplus* of possible states subsystems A and B can assume alone but summed up, relative to the states that their combination may be in. If A and B are two neurons that can each assume two possible states (0 or 1, inactive or active, with 50% probability), their individual entropy ($H(A) = H(B)$) equals 1 (see equation 5.1). If these neurons behave independently, the pair can assume four states (A on and B off, A on and B on, etc; see figure 5.5), so its entropy is $^2\log 4 = 2$. With statistical independence the mutual information is thus $1 + 1 - 2 = 0$. But if neuron A and B are fully coupled, so that B is always active when A is active and B turns inactive whenever A is silenced, the system has only two degrees of freedom ($H(AB) = ^2\log 2 = 1$). Mutual information reflects this logic and becomes 1. A high MI value can be achieved under two conditions: (1) when statistical dependence between system elements is high and (2) when subsystems A and B can assume a large number of different states (Papoulis, 1991). If the system is extended to include more neurons, equation 5.4 can be rewritten as follows:

$$\text{Int}(X) = \sum H(x_i) - H(X), \quad (5.5)$$

where an individual neuron x_i is a member of a neuronal group, X , and $\text{Int}(X)$ defines the integration within X . To tease out whether a group of neurons X constitutes a distinct cluster within a larger system of neurons, a cluster index CI is proposed,

$$\text{CI}(X) = \text{Int}(X)/\text{MI}(X;Y), \quad (5.6)$$

which is the ratio between the statistical dependence within group X relative to the statistical dependence between X and the rest of the system Y . Tononi (2004) used somewhat different measures based on mutual information, but these differences are less relevant here.

To determine how *differentiated* a system is, Tononi and Edelman considered not just a single subset of neurons but all possible subsets a large neural system can be composed of. They reasoned that the differentiation of conscious experience is best specified by the number of different states a system could possibly assume. The neural complexity (C) of a larger system X is defined as follows:

$$C(X) = 1/2\Sigma < \text{MI}(X_j^k; X - X_j^k) > \quad (5.7)$$

where X_j^k is the j -th subset of X and k indexes the number of elements in each subset. $(X - X_j^k)$ is the complement of X_j^k and thus represents the rest of system X . A summation is carried out over all possible subsets of X , starting with each individual neuron within

X ($k = 1$) against all the rest, going on with pairs of neurons ($k = 2$), and so forth. The average mutual information MI is computed across all j th combinations of k elements, as indicated by the $\langle \dots \rangle$ symbols. Applying such measures to cortical dynamics in simulated neural systems as well as EEG recordings of sleep–wake states and epileptic seizures, Tononi, Edelman, and colleagues confirmed their usefulness in describing neural group dynamics (Tononi et al., 1994; Alkire et al., 2008).

The overall effort to develop a quantitative, axiomatic approach to consciousness should be applauded. Nonetheless we need to ask what it fundamentally clarifies about the neural basis of consciousness. The approach is to quantify integration by detecting enhanced statistical dependence in a subset of neurons relative to their peers not part of that subset. This is less specific than the *functional* and *psychological* concept of integration Tononi and Edelman introduced earlier on in their 1998 paper:

integration is a property shared by every conscious experience irrespective of its specific content: each conscious state comprises a single “scene” that cannot be decomposed into independent components. (p. 1846).

Indeed “scene construction” can be reasonably considered a core property of consciousness. However, these informational measures of integration do not show that, or how, consciously experienced scenes are constructed. A variety of processes are needed, including binding individual scene elements or attributes into single objects (see chapter 3), ordering objects in the scene according to spatial positioning, and conveying phenomenal content to objects and properties, so they come to refer to properties the brain constructs these objects to have in the scene setting. These are different, although co-occurring and intertwined, processes to be discussed later. The point here is that the proposed measure of informational integration does not add explanatory power as to how these processes may work. This is not to dismiss the approach: it is highly valuable to develop quantitative measures of integration and differentiation in their own right. Development of further quantitative approaches addressing the validity of other proposals (e.g., Seth et al., 2006) is to be encouraged.

What about complexity as a measure for how “rich” consciousness experientially is? Experiences can be set in multiple modalities, and scenes seem to be incredibly rich in their qualitative details (although it can be debated whether we are simultaneously conscious of all such details). But is complexity *per se* a defining characteristic of consciousness? If we turn away from our usual fixation on vision and focus on a sensation such as the intense smell of ammonia, do we then experience something very *complex* or rather something all-at-once, simple and unified (cf. Hopfield, 1994)? Two different notions of complexity are at stake here. First, a conscious experience of smelling NH_3 arises as one out of many possible alternatives, but this holds for many multielement systems throughout nature (the panpsychic argument). Second, the experience itself may vary in complexity, depending on the variety and quantity of objects and properties present in the scene we experience.

Conscious experience does not have to be complex. We can simply taste fresh lemon ice cream, closing our eyes and oblivious for a moment about our surroundings, past and future. No matter how complex the *neural mechanisms* underlying consciousness might be, we should not have our models of conscious representation generate *experiential complexity* for the sake of complexity.

Further challenges arise for the quantification of conscious experience based on the proposed information-theoretical measures. First, differentiation and integration are gauged in a rather static manner, requiring averaging across considerable laps of time to obtain robust statistics. For testing in practice, it would be mandatory to use measures that can be deployed on a single-trial basis, so that one may test whether a repeated trial, with the same psychophysical report, consistently results in more similar values than when percepts change. Another problem arises from the specification of neural complexity as a collection of a system's *possible* states, not just the one that a subject is currently experiencing. Tononi (2004) characterizes consciousness as a disposition or *potentiality*—"in this case as the potential differentiation of a system's responses to all possible perturbations, yet it is undeniably *actual*" (p. 19).

However, this paradox between consciousness as a potentiality or actual state is not resolved. If a conscious state is defined as a differentiated and integrated whole, this should be taken to refer to the rich content the experienced scene *currently* has, not a disposition the system may have to generate many different states in the future. In contrast to a dispositional view of consciousness, Tononi and Edelman took (actual) *neural processes* as the starting point of their 1998 paper, which is also the view I will adhere to here. It is undeniably so that at any given time a conscious system assumes only one state out of an enormous multitude of states it *could have* assumed at that moment, but this does not render the phenomenon a potentiality. The specificity of conscious experience relates to a system's actual state, not all other states possible in the past or future. Thus we should seek further measures to compute the "richness" or differentiation of a scene based on its actual content.

Revisiting panpsychism, many multielement, dynamic systems throughout nature will percolate through many corners of their state space. Water particles in a tropical cyclone can be said, like neurons, to be sensitive to external ("sensory") inputs such as air pressure, airflow, temperature, and light. The statistical dependence of their behavior will be strongly dependent on these inputs—for instance, when a drop in temperature causes condensation or a sharp gradient in ambient air pressure causes coherent movement. If this metaphor is criticized for its lack of functional clustering behavior, think of a glacier that starts melting under the influence of sunlight in spring. In the crystalline structure of ice, water has low entropy, and despite high statistical dependence the complexity of ice will be low (comparable to slow-wave sleep or epilepsy). Once the ice melts, molecules gain more degrees of freedom, but their movements remain statistically coupled to their neighbors, hence raising complexity (like an awake neural system with high-density connections). Functional clusters in this system are identified as pockets in which the phase transition between ice and

fluid occurs earlier than elsewhere, inhomogeneously spread across the melting glacier. It is straightforward to identify further examples of inanimate systems showing a high degree of “complexity” and “integration.”

5.6 Microconsciousness or Atomic Consciousness

Informational measures to quantify statistical relationships between neuronal or other physical elements concern the *syntax* by which a system operates. Syntax is not about the semantics or contents of representations. It is not surprising that information theory has offered little so far to elucidate mind–brain relationships, because semantics does not fall within the scope of its intended use (Shannon, 1948, 1949) and the genesis of meaning has not been among its targets.

We may also conclude that, until now, models for conscious representation are subject to the criticism that systems of comparable complexity can be found all too easily throughout nature, without a convincing link to consciousness. An alternative philosophical view is that of “atomic consciousness” or “atomic panpsychism,” which views consciousness not as an emergent property but as a fundamental quantity of system elements, like the electric charge of elementary particles.

For instance, Penrose and Hameroff applied ideas from quantum mechanics to consciousness research, theorizing that certain cytoskeletal elements in neurons (microtubules) may link quantum behavior at the atomic level to neuronal activity and hence consciousness (e.g., Woolf & Hameroff, 2001). Chalmers (1995) suggested that conscious experience should be thought of as an elementary physical quantity, comparable to mass, charge, or space–time. Also in neuroscience, Zeki and Bartels (1999; Zeki, 2008), Tononi (2004), and Koch (2012) argued in favor of “microconsciousness,” panpsychism, or similar positions. I will argue that this line of reasoning is understandable, but not helpful or necessary.

Single neurons, or even single molecules and atoms, can be said to have the *potential* to contribute to consciousness, especially when placed in a proper configuration interacting with other system elements. This potential is evidenced by the fact that our own bodies are made of molecules and atoms, and we are conscious for most of the day. But to attribute an “atomic consciousness” to particles, or a “microconsciousness” to single neurons, is a different matter. A first problem is the lack of evidence that certain properties of elementary particles *directly* contribute to consciousness. We do have evidence for some linkage between electromagnetic forces and consciousness, but this relationship is very indirect, running via phenomena such as spikes and neural network activity to macroscopic phenomena.

Second, even if we were to discover a new fundamental force or particle that would endow neurons with consciousness—let’s call it the C-force for convenience—a whole new theory and experimental program would be needed to explain a *possible* link between *this* force and consciousness. We would have to explain how the force gives rise to signals

that code sensory content, motor decisions, and every other process we associate with cognition (cf. Dennett, 2005). We would *also* be compelled to link this secretive force to bioelectricity, which we *can* already correlate to cognition and consciousness.

Supposing a C-force exists, it would be unclear how “conscious” neurons, sitting in our brains, would be fed sensory content. Reflected by a palm tree on a sunny beach, patterns of light fall on your retina and activate relevant groups of neurons, which would then switch from a conscious-but-void state to a conscious-of-palm-tree state—but how would this content switch happen? Do transmitters and ion channels still play a role? If not, other unknown factors should be at play, widening the net of implausible assumptions. If yes, how should we suppose that an influx of Na^+ ions feeds sensory content into an already conscious neuron? And if a neuron’s experiential content has any effect on your actions—such as unfolding your towel on the soft sand—how would a conscious neuron convey its information toward your musculature? We cannot fully exclude the existence of a C-force. However, in the absence of evidence, this research direction would be superfluous in the face of the vast body of evidence linking electrophysiology to cognition.

A third problem relates to the pragmatic use of the concept of consciousness. If we were to adopt consciousness as a fundamental property of nature, then every object and particle we encounter would have mind-like properties. We would be free to discard any conscious–unconscious dichotomy. But is it practically useful to disregard this distinction? If not by terms like consciousness, how else should we describe differences in brain and behavioral states such as wakefulness versus sleep or coma, or being aware of a traffic sign versus not having seen it?

When evaluating the plausibility of microconsciousness or consciousness as a new, fundamental physical property, my plea comes down to empiricism. When a dentist’s syringe squeezes lidocaine on my trigeminal nerve, the toothache vanishes—and it makes little sense to ascribe the molecular action of this anesthetic to anything other than Na^+ channel blockade in nerve fibers. Electrical microstimulation of motion-sensitive cells in a cortical area of the macaque’s visual system can influence the monkey’s perceptual judgment of the average direction in which a cloud of dots on a visual display screen is moving (Salzman et al., 1990). No need arises to explain this effect by physical phenomena other than the effect of electrical pulses applied via an electrode on the firing activity of nearby neurons and axons. Examples of correlations between electrophysiological activity and perceptual phenomena are so abundant (e.g., Leopold & Logothetis, 1999; Supèr et al., 2001b; Kreiman et al., 2002; Del Cul et al., 2007; Fisch et al., 2009) that, in the absence of positive evidence, the assumption of novel physical phenomena pulling off the trick appears redundant and unnecessary. In conclusion, there are solid reasons to reject atomic and microvariants of consciousness, associated with panpsychism. A better, panpsychism-resistant theory is needed of what conscious representations are about—of how the brain builds an integrated, ordered, meaningful scene or experience.

A second conclusion is about the usefulness of nonliving complex systems as a kind of complementary Turing test. Turing's (1950) idea was to consider an artifact as "intelligent" only if a human observer is unable to distinguish whether he is communicating with this artifact or another, human agent, unaware of the physical appearance of both. The variant proposed here is that a model or artifact for conscious representation is only valid as such if it is not reducible to a system that is *patently incapable* of passing a Turing-type examination of its behavior, anatomy, and physiology. In this test we are not targeting intelligence but conscious experience, as exemplified by pain, vision of scenes, body movement, food and wine tasting, and so forth. We may begin with behavior as a first litmus test, citing Wittgenstein (1953): "if I see someone writhing in pain with evident cause I do not think: all the same, his feelings are hidden for me" (p. 223). However, because behavioral tests, as stand-alone indicators, are not foolproof, it is critical to devise supplementary anatomical and neurophysiological criteria. In the case of conscious pain, for instance, we should aim to identify a complete set of criteria to establish the behavioral characteristics of conscious pain, of anatomical and connectional architectures involved, as well as of the physiology capable of supporting multimodally varied and specific pain representations. Models resembling rock networks or melting glaciers do not pass this "panpsychism" test, but there is no principled reason why more advanced models should invariably lead to rejection. The test may clarify why and how current theories are underconstrained as models of consciousness and how they can be extended.

Thus, a panpsychism-resistance test can be used for a first triage by inferring whether models can be reduced to systems that cannot be reasonably qualified as conscious—not only failing to emit overt behavioral signs but also lacking sufficient representational capacities. In addition, we may identify better models of consciousness by showing that their built-in mechanisms (e.g., neural connections, interactions) are sufficient to account for the representational capacities we associate with conscious experiences.

6

Structure and Function of Brain Systems for Conscious and Nonconscious Representation

Although *given time* a suitably programmed universal machine can perform any specified task, this gives no guarantee that we could get any such machine, however large, to simulate all the multiple interactions in a brain *simultaneously*, and in *real time* [...]. There is enough evidence in this general area to suggest that we on the circuit side had better be very cautious before we insist that the kind of information processing that a brain does can be replicated in a realisable circuit. Some kind of “wet” engineering may turn out to be inevitable.

—D. M. MacKay (1965)

6.1 The Biological Function of Consciousness

If models for representation, either of a neural-network or psychosemantic variety, are unable to tell us why we experience some pieces of information in our brain as conscious whereas others remain nonconscious, what could we learn from the structure and functioning of brain regions themselves? Here I will not offer a comprehensive overview of all brain structures but will focus on a few systems that are especially important from a representational viewpoint: corticothalamic systems, cerebellum, basal ganglia, and hypothalamus. By far, cortical and thalamic structures have received most of the attention in studies of consciousness, but the rationale to put the other systems under the magnifying glass is that they have *not* been associated with consciousness in most of the literature. What would make the cortex and thalamus so special? Which structural and physiological properties might be lacking in subcortical structures, denying them a participation in the feast of awareness? Can we identify functional motifs as to why these structures would be “condemned” to nonconscious processing? Especially this last issue deals with basic functions of different brain systems for survival and reproduction, which begs the question what kind of function consciousness, if any, may have in the first place.

Darwin taught us to think of organisms as surviving evolutionary combat by coping with stimuli that either threaten them or promise to satisfy one of their needs, such as a predator, prey, water, shelter, or sex. When we track a tadpole swimming through a pond, we observe how she supplies herself with fresh, oxygenated water by spontaneous tail movements. While exploring the environment, she will be guided by light input to aquatic plants she

can graze on. Her nasal chemosensors might detect a gradient of some harmful contaminant, driving her in the opposite direction. A dark, beak-shaped shadow falling across the water's surface causes her to duck away.

Individual stimulus-response associations suit the organism's survival perfectly well, but how would it be if a behavioral decision has to be based on a complex of stimuli offered all at once? Much more food might be found in an environment *A* that resembles an environment *B* in a few salient stimuli but differs in a number of details, such as how local duckweed strings are configured in space. It is adaptive not only to respond to singular stimuli by individual actions but also to represent information about many stimuli present in the environment *all at once*, so a behavioral decision can be based on this complex whole. One basic strategy of nervous systems throughout evolution has been to develop sensors for dealing with individual stimuli in different modalities. A more advanced strategy is to construct an integrated, contextual representation of the actual situation in the world and the organism's own state, allowing more refined and anticipatory decision making. Stimuli can conflict with each other in that they would evoke opposite responses when dealt with individually. Often different sensory inputs have to be combined for optimal decisions to be taken (only if the shape of an apple is combined with "red" should I bite, not when "brown"). If you play soccer and find yourself with the ball at your feet in front of a goal, you have to determine in a split second whether you are at your own team's goal or the opponent's—but you cannot base this decision on the features of the goal alone.

Thus our brains have to integrate all incoming pieces of sensory information into a coherent representation of the situation that must be acted upon, and they have to do so very quickly, as environmental threats may strike immediately and opportunities may vanish into the brushwood. This situational representation should be as detailed in content as a short time frame allows our brains to construct, and a key aspect of representation is that it is fundamentally *multimodal* in nature—it integrates information from our main modalities (vision, hearing, touch, smell, taste, proprioception, vestibular sense, pain, and thermoception). Dominant as any modality may be at a given moment, we can see, hear, feel, smell, and taste at the same time (which, by itself, does not suffice to speak of "integration").

The task of rapidly constructing a multimodal, situational representation capturing at least roughly what is going on in the world is what I call the brain's representational problem. The brain's solution to it is what we call consciousness or conscious representation. This concept is not entirely new. For instance, Koch (2004) proposed that visual awareness presents us with the "gist" of a visual scene—a succinct summary of "what is in front of me" (see also Crick & Koch, 2003). Earlier on, contemporary philosophers already emphasized the singular, unified nature of conscious experience and its characteristic first-person perspective (e.g., Searle, 1992; Churchland, 1995; Metzinger, 2000). Also Marr's (1982) model on the brain's construction of progressive "sketches" for vision (primal, 2.5-D, and 3-D) incorporated the notion of a spatial representation for conscious vision. In the end, these and other related concepts all pay tribute to Kant's original notion

of space, put forward in his *Kritik der reinen Vernunft* (1787), as an a priori form of appearance in which all experience is set.

The current account broadens consciousness to encompass all modalities. I prefer to speak of representations of *situations* instead of “scenes,” which are strongly connotated to vision. Moreover, conscious representations need not to be limited to the gist of a scene (an “executive summary” needed for quick decisions) but can be detailed as well—depending on processes like foveation and attention. If you are standing in front of Caravaggio’s *The Calling of St. Matthew* (see figure 6.1, plate 2) in the San Luigi dei Francesi in Rome, you experience not only “what is in front of you” but also your body posture (including your own back, which you cannot see), the ambient lighting conditions and baroque architecture, the whispering of other visitors, and the sound of Gregorian chants. You can be captivated by the painting’s visual qualities while other modalities linger at the fringe of consciousness. Probably you are not aware of all details of the painting at once, even if you have the feeling that you are (Kouider et al., 2010, 2012). However, if your feeling of weight (gravity), sense of balance, and peripheral vision would be deleted from your experience, it would change dramatically. It would lose a great deal of its sense of “immersion.” Experiences can be rough and flash-like, as when you are driving a car and you recognize somebody passing in the opposite direction, but they may also be extended and detailed, as when you scrutinize the light beam illuminating St. Matthew.

Does a situational representation have to contain all of the *qualities* we typically associate with conscious perception? Why not replace our brains with a high-resolution camera and a bunch of body-position and motion sensors and send all of their outputs into a huge working memory buffer? Doesn’t this memory represent the same kind and amount of information? The argument here is that the brain’s neurons are way too slow for the organism’s needs to represent sensory information by long strings of zeros and ones, as a conventional personal computer would do. A major problem with understanding the input “000001100110101001010110”—regardless of whether this represents electromagnetic states on a chip or a neuron’s spike train—is that it leaves unsolved what *category* the information belongs to, how it can be *partitioned* into different elements such as objects, and how to *order* it according to spatial location, time, or properties such as common motion. These are integrative tasks the brain should be able to carry out immediately, and almost continuously during the day, as long as we are awake.

The most fundamental step in categorizing information is to identify the sensory modality a piece of information belongs to, and this step can be considered the primordial stage of grasping what an input “means” or “feels like” at a very raw, sensory level. Further steps include the classification of objects as concepts we have learned to recognize later in our development, such as a duck or a rabbit’s head (see figure 4.5). This is already a step toward assigning *conceptual* meaning to an object, which allows us to associate it with related concepts.



Figure 6.1 (plate 2)

The Calling of St. Matthew by Caravaggio. This painting and its setting in the San Luigi dei Francesi church in Rome illustrate the concept of conscious experience as a dynamic, multimodal, situational representation.

In short, the brain needs integrative neural mechanisms to accomplish a multimodal situational representation, not only to *detect* any particular piece of sensory input in it but to represent it together with other inputs in the fore- or background and to render it meaningful within the context of all other information received. A purely numerical code would be unfit for this task, at least at the higher level the brain is working at to organize perception. Instead, I submit the solution to the representational problem is that the brain generates an internal model rich in qualitative properties to make information immediately graspable and set within a spatiotemporal framework. The more kinds of qualitative property a system can integrate in its representations, the higher its representational capacity and the “richer” the potential for conscious experience. This reasoning yields a clear (functional, teleological) reason *why* conscious experience is loaded with qualitative properties, whereas an explanation of how the brain pulls off the trick remains to be given.

The same reasoning provides a rationale for the brain’s also having systems for nonconscious information processing. We encounter many daily-life situations in which no high-level representations are needed because adequate sensorimotor solutions can be found by processing singular stimuli. In reaction to a rush of cold air, your arm hairs stand up in a reflex. If all singular stimuli were included in conscious representations, they would

become overloaded. It would take too much time to generate a situational overview, and this overview would become too crowded to enable efficient motor and cognitive decisions. This leads us to relate conscious processing to the speed and timing of decision making (cf. Libet et al., 1983). On the one hand, we need “ultrafast” action selection such as for quickly fencing ourselves off from a fast approaching object such as a bird flying toward us. Here it is safer to rely on reflexes and automated stimulus–response reactions before a conscious representation has been fully assembled. On the other hand, decisions requiring a situational overview should still allow decision making in a medium-fast range, arguing that only those elements of a situation should be consciously represented that are salient and cannot be dealt with the “easy way,” that is, automatically.

Once we have a good situational update of the world around us and of ourselves, we usually do not need another elaborate experience of the *motor solution* to respond to the situation at hand. A conscious overview is needed to grasp Caravaggio’s masterpiece, but we are better off without a conscious representation of how our eye muscles implement tracking movements, or of eye movement itself. Motor commands and even higher planning operations can be carried out in an automatized, nonconscious fashion. We can easily become conscious of action *consequences* as these lead to altered proprioceptive, vestibular, and somatosensory feedback from body to brain, but this is different from being aware of individual motor commands or muscle contractions.

We may also consciously *think* about actions we may undertake, but those aspects of thought are characterized by their specific (quasi-)sensory nature. First, we can be conscious of thoughts generated by visual *imagery*. Second, thinking can be instantiated by *auditory imagery* and associated *self-generated sensations* during subvocal articulation of words (“speaking to oneself”). Thus, the functional account of consciousness I am proposing leaves plenty of space for nonconscious information processing to exist. This functional—but not necessarily functionalist—account implies that consciousness does not constitute an epiphenomenon, a by-product of other processes doing the real work. Once your brain integrates information into the right sort of representation, you have full-blown, functional consciousness.

6.2 Visual Corticothalamic Systems: Local Circuits

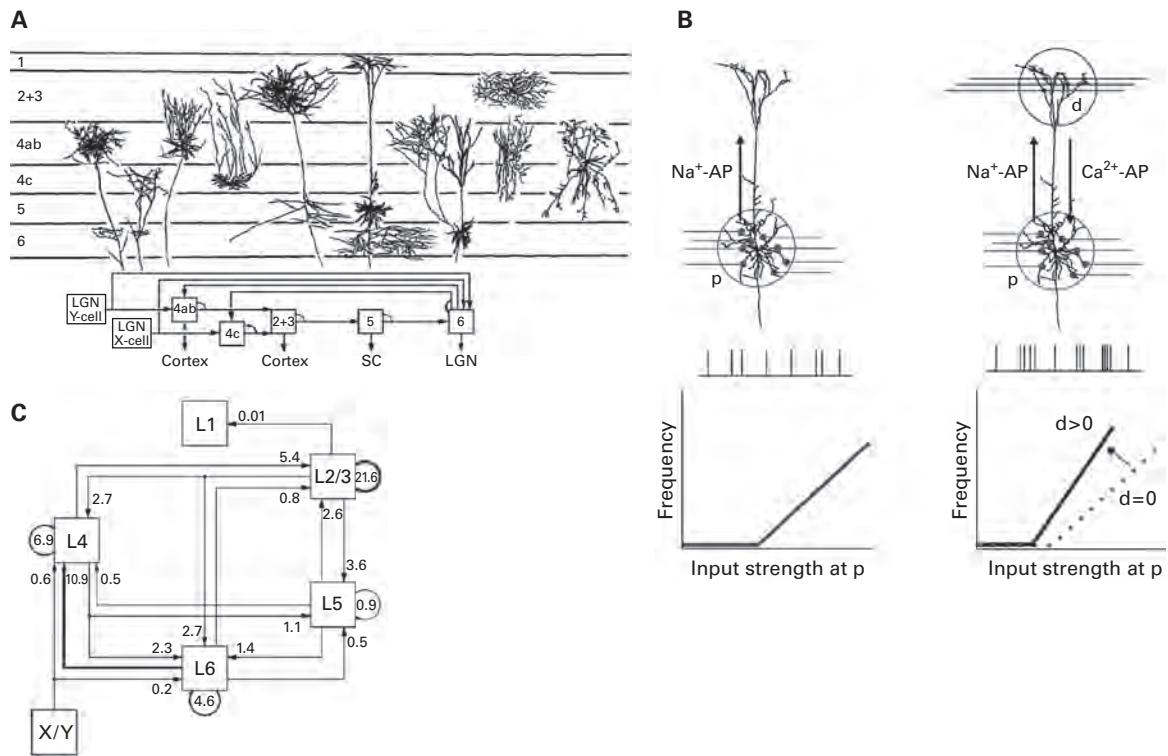
Which brain systems qualify to construct multimodal, situational representations? Chapter 3 pointed to primary and higher-order sensory cortical areas as core components. Here we will examine which structural features of these areas and their thalamic counterparts may be important for realizing this function. Whereas the vast majority of consciousness research has focused on vision, we will also pay attention to somatosensation, audition, and olfaction. If a given component thought to be crucial for visual awareness is lacking in a different cortical system, that component should not be considered critical for consciousness in general.

In chapter 2 we saw how the cortex is composed of six layers, stacked on top of each other like pancakes. Cortical areas are classified as having a layer 4 (L4) or as lacking this layer. By and large, areas harboring a L4 are sensory in nature, whereas those without one are located mostly in the frontal lobe but also include some temporal lobe areas (Rempel-Clower & Barbas, 2000; Wise, 2008; Shepherd, 2009). Because L4 has a granular (grainy) appearance due to its numerous stellate (“star-like”) neurons, subsets of frontal areas are also referred to as “agranular” or “dysgranular.”

When we try to track the flow of sensory information in a local wiring diagram of visual cortex based on classic work in monkey studies (Lund, 1988), the L4 stellate cells act as main recipients of inputs from the thalamic relay nuclei. Their outputs are emitted to neurons in L2–3, mostly within the same cortical microregion (see figure 6.2A). In primate visual, auditory, and somatosensory cortex, “microregion” can often be translated as “column,” that is, a vertically organized anatomical entity having a stereotyped intrinsic wiring and common functionality of the neurons it contains (Powell & Mountcastle, 1959; Hubel & Wiesel, 1977; Linden & Schreiner, 2003; Douglas & Martin, 2007). However, columns have not been discretely recognized in many other cortical systems or even species.

L2–3 pyramidal cells relay spike information to other cortical areas, and in addition a portion of L4 cells do so. In primary visual cortex, much of the corticocortical connections are made with “higher,” extrastriate areas (V2, 3, 4, 5, and MT). Another portion of these connections is made with L2–3 cells in the same area (McGuire et al., 1991), thereby labeled “horizontal.” Horizontal axons tend to connect neurons having the same functional specificity, as illustrated by their response preferences for visual stimuli with similar orientation (Ts’o et al., 1986; Ko et al., 2011). The large pyramidal cells which have their somata in L5 stick out their apical tuft dendrites into the superficial layers, ramifying into finer branches (see figure 6.2B). These large cells receive inputs from local superficial cells but also pick up signals from more remote cortical areas. In turn, L5 cells in visual cortex emit signals to subcortical areas such as the superior colliculus and pulvinar (a large nucleus positioned caudally in the thalamus, considered “higher-order” because it does not receive direct input from the retina but is fed visual cortical information). In addition, L5 cells project further down to local L6 neurons, which in turn send output to local L4 cells or project outside the cortex—for example, back to the thalamic relay nucleus that supplies the same cortical area with input (cf. figure 2.5). Thus, L6 cells are thought to be at the core of a corticothalamic feedback loop, with the addition that the thalamic reticular nucleus, fed with L6 output, provides inhibition on thalamic relay neurons (see figure 2.5). So far, these circuit connections are thought to be excitatory and use glutamate as transmitter.

In this way, various “loops” or closed, excitatory circuits might be formed, in particular the following: L4 → L2–3 → L5 → L6 → L4 (see figure 6.2A). However, work that took place in cat and rodent visual cortex since the classical studies on monkey V1 identified many other loops, particularly because many laminar connections are reciprocal

**Figure 6.2**

Classic and contemporary schemes of information processing in neocortical circuits. (A) Using intracellular injections and anatomical reconstruction of recorded and labeled neurons distributed across layers of primary visual cortex, Gilbert, Wiesel, and colleagues proposed a circuit scheme for cat area 17 in which layer 4 cells, activated by lateral geniculate nucleus (LGN) inputs, predominantly activate a sequence of layer 2–3, layer 5, and layer 6 cells (alternative routes are also indicated here). Abbreviations: SC, superior colliculus. X- and Y-cells in the LGN refer to cells with different dynamics of responding to the on- and offset of visual stimuli. (B) A layer 5 cell is shown in two different situations: on the left, the cell receives excitatory inputs via its basal dendrites in and around layer 5 (also called “proximal,” p), which triggers a Na^+ channel-mediated spike propagated along its axon but also along the apical dendrite penetrating into the superficial layers. On the right, additional inputs in the superficial layer (also called “distal,” d) trigger a Ca^{2+} channel-mediated action potential that is propagated from the apical dendrite into the soma. The result of this interaction is that the input–output relationship becomes steeper (lower panels) and the threshold for spiking (expressed in frequency or firing rate) is lowered across the same range of proximal input strengths. The temporal spike patterns (middle panels) also become more bursty as a consequence of the Ca^{2+} action potential. (C) Contemporary circuit diagram of cat area 17, quantitatively estimating the strength of excitatory connections between local neurons and thalamic afferents. This analysis was based on elaborate axonal and dendritic reconstructions of single, intracellularly recorded and stained neurons. Note how the classical circuit in (A) is complemented with many more feedforward and feedback connections, the strength of which is symbolized by line thickness with accompanying weights. (A) from Gilbert (1983), with permission from Annual Reviews Inc. (B) from Larkum et al. (2004), with permission from Oxford University Press. (C) from Douglas and Martin (2007).

(Binzegger et al., 2004). For instance, L4 not only emits projections to L2–3 but also receives major inputs back from these layers (see figure 6.2C).

Until here, inhibitory interneurons of the cortex have been left out of the picture (Ascoli et al., 2008; Gentet, 2012). Without them, cortical circuits would be at risk to become hyperactive, as positive feedback loops would cause firing activity to run up to maximal levels, exhausting themselves in an epilepsy-like state. In addition, GABAergic neurons are recognized as crucial elements for pacing thalamocortical circuits into rhythmic mass activity and may fulfill specific computational functions in converting inputs to a cortical microregion into its outputs.

Modern electrophysiological and optogenetic techniques have brought further updates on the wiring diagrams derived from anatomic tracer studies (e.g., Yizhar et al., 2011; Petreanu et al., 2007, 2009; Lefort et al., 2009; Barbour & Callaway, 2008). For instance, patch-clamp studies in multiple cortical areas have shown that L5 pyramidal cells receive excitatory input not only via the superficial layers but also via their basal dendrites from other neurons in the deep layers (Larkum et al., 2004; Bannister, 2005). This addition is more than a refinement because it presents deep-layer cells not only as “output” units of the neocortex (see figure 6.2A) but as central processing elements, integrating and associating inputs coming in via both deep and superficial layers (Feldmeyer & Sakmann, 2000).

In summary, the classic loop scheme of figure 6.2A has been replaced by a much more intricate, densely connected circuit diagram of visual cortex, making it more appropriate now to consider columns or microregions as intricate, dynamic networks in their own right.

6.3 Visual Cortex: Ascending and Descending Projections

Are such local circuits homogeneously repeated across all cortical areas, or do these areas have different connectional relationships with one another? Rockland and Pandya (1979; Rockland & Van Hoesen, 1994) approached this question by classifying cortical area-area connections as either “ascending” (or feedforward) or “descending” (or recurrent, feedback), depending on the layer in which axons of the “sending” area terminate. If a sending area emits a projection from its L2–3 terminating primarily in L4 of a target area, the projection is labeled “ascending,” whereas a termination mainly in the superficial layers (including L1, but sometimes also L6) is “descending.”

Based on these topological rules, Felleman and Van Essen (1991) proposed a well-known hierarchical scheme of visual system organization in primates. For instance, area V2 occupies a relatively “low” position because its L4 receives cortical inputs only from V1, while V2 receives recurrent inputs from higher visual areas, but also from prefrontal regions and the frontal eye fields (see figure 6.3, plate 3), which exert control over eye movements. Usually the relationship between two given areas of cortex is reciprocal. In the case of V2, the ascending projection from V1 is reciprocated by a descending projection to V1, reaching its superficial layers. In addition to the hierarchical (“vertical”)

relationships between areas, the scheme acknowledges many “lateral” or “horizontal” connections. These point to areas that are anatomically connected but have no clear hierarchical relationship of ascending–descending projections to one another. Felleman and Van Essen pointed out that there are many irregularities in the scheme, and it is somewhat unclear whether these may be resolved by applying additional connection rules or whether the visual system is quasi-hierarchical at best.

What could be the function of this hierarchical organization? This question is difficult to answer at this time. Perhaps it permits an orderly flow and hierarchical integration of information from primary to higher visual areas, from low-level processing of local features in the visual field to the high-level, wide-field processing of complex objects such as faces and cars, as occurs in the inferior temporal cortex (Riesenhuber & Poggio, 1999; Tanaka, 2003; Rolls, 2012). Consistent with this, the hippocampus sits at the apex of the hierarchy, where cells are sensitive to complex spatial configurations of stimuli (O’Keefe & Dostrovsky, 1971; McNaughton et al., 2006).

Sensory input entering the cortex can be thought of as being propagated by a “feed-forward” wave traveling from primary to higher sensory areas (Lamme & Roelfsema, 2000; VanRullen & Thorpe, 2002)—with the important qualification that the input does not remain constant but is transformed along the way. Once this wave of activity reaches a higher area, a backward, “recurrent” wave may arise. This wave may then feed highly specific information on objects and contexts back to primary areas. At the same time, higher-level areas engage not only in sensory processing but also in long-term and working memory, categorization, emotional valuation of stimuli, planning, and decision making. Hence, recurrent pathways may carry the effects of such cognitive operations back to lower areas (e.g., Supèr et al., 2001a; Shuler & Bear, 2006).

One problem in this account is that it does not explain the very function of a circuitry that carefully distinguishes between ascending versus descending information. One might propose that the distinction is useful for keeping feedforward information trafficking across the cortex segregated from feedback processing: the L2–3 projections from a lower area *A* to L4 in a higher area *B* channel feedforward processing, whereas the L2–3 back projections from *B* reach other layers in the area *A* (viz., L2–3 and sometimes L6). However, this segregational account does not hold up because the very superficial layers of *A*, receiving the feedback, emit forward projections to *B*. Thus, it is not yet clear what advantage the ascending–descending hierarchy would bring over a scheme where all corticocortical trafficking would be mediated by superficial connections only.

Regardless of this, L4 stellate cells may fulfill a specialized role in processing bottom-up inputs, which would not be needed when processing recurrent inputs. Higher visual areas lack a direct input from the visual relay nucleus in the thalamus (LGN), and so the feedforward input they receive from lower areas may act as a “thalamic input by proxy,” requiring the same kind of computational operations as needed to process LGN input reaching V1. Considering that thalamic inputs synapsing on stellate cells are relatively sparse, Douglas

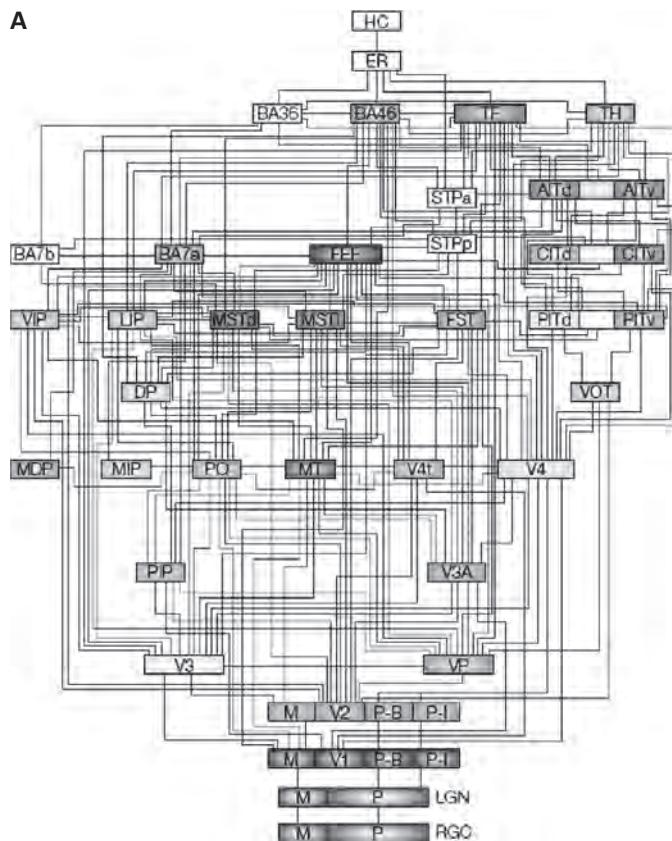
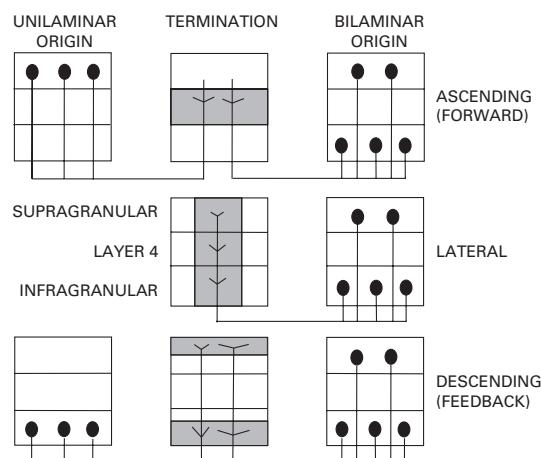
A**B**

Figure 6.3 (plate 3)

(A) A hierarchy of visual areas as proposed by Felleman and Van Essen (1991), based on a compilation of anatomic tracer studies. Starting at the level of retinal ganglion cells (RGCs; bottom) and lateral geniculate nucleus (LGN; M and P denote magno- and parvocellular pathways), the projections conveying photic information spread out across a network of 32 cortical areas. A total of 187 area-area connections are shown, most of which are reciprocal. Besides up- or downward projections, many areas are connected laterally. Note how the entorhinal cortex (ER) and hippocampus (HC) involved in declarative memory are located at the apex of the hierarchy. (B) Whether an area is placed at a higher or lower level depends on its pattern of afferent and efferent connections across its laminae. All boxes shown are divided into three zones: (1) supragranular or superficial layers, (2) layer 4, and (3) infragranular or deep layers (middle panel). If a given cortical area sends a projection from its superficial layers to layer 4 of a receiving area, the projection is considered ascending or forward. The same applies if both the superficial and deep layers of the sending area are projecting. For descending projections, terminations in the superficial and deep layers of the target area are critical. Terminations of lateral connections are distributed across all three laminar compartments. Reprinted with permission from Oxford University Press.

and Martin (2007) proposed a filtering and amplifying role of L4 circuitry. By firing only to a small, optimal subset of all inputs, L4 cells may respond very selectively to visual features, and their local reciprocal connections with cells in L2–3 and L4 may amplify this selective response. This proposal awaits experimental testing but appears consistent with a recent optogenetic study in mouse auditory cortex, suggesting a role for intracortical excitation in amplifying thalamocortical responses (Li et al., 2013).

One of the caveats on figure 6.3 (plate 3) is that probably prefrontal, temporal, and parietal regions do not obey the connectional hierarchy strictly as defined above by differential laminar projections (Webster et al., 1994; Rempel-Clower & Barbas, 2000; Medalla & Barbas, 2010). Furthermore, the criteria Felleman and Van Essen used to compose their hierarchy may be underconstrained, leaving space for other optimal orderings. This has raised the suggestion that the organization of cortex may be inherently indeterminate (Hilgetag et al., 1996; Reid et al., 2009).

Figure 6.3 (plate 3) almost makes us forget that information is also communicated via corticothalamic circuits, with L5–6 cells as critical relays. For instance, V1 receives its visual input mainly from LGN and projects back to the LGN but also to the pulvinar by way of L5 cells. The pulvinar projects, among others, to all extrastriate visual cortical areas and parietal area 7 (Sherman, 2006). The auditory and somatosensory systems contain similar, higher-order thalamic relays, namely, the magnocellular portion of the medial geniculate nucleus and the medial part of the posterior thalamic complex, respectively (Sherman, 2006; Theyel et al., 2010). Thus, the flow of information across the neocortex will be interrupted or sculpted by routing via the higher-order thalamic nuclei. The situation is further complicated by the surrounding reticular nucleus, boosting different kinds of mass oscillations in lower- and higher-order cortical areas (Steriade, McCormick, & Sejnowski, 1993; Crunelli & Hughes, 2010; Suffczynski et al., 2001).

A final remark is that “top-down” feedback is not restricted here to those cortical regions that are usually associated with psychological top-down processing. In psychology, top-down processing refers to cognitive (knowledge-driven) control over low-level processes such as sensory analysis. This top-down control is thought to be exerted by higher cortical

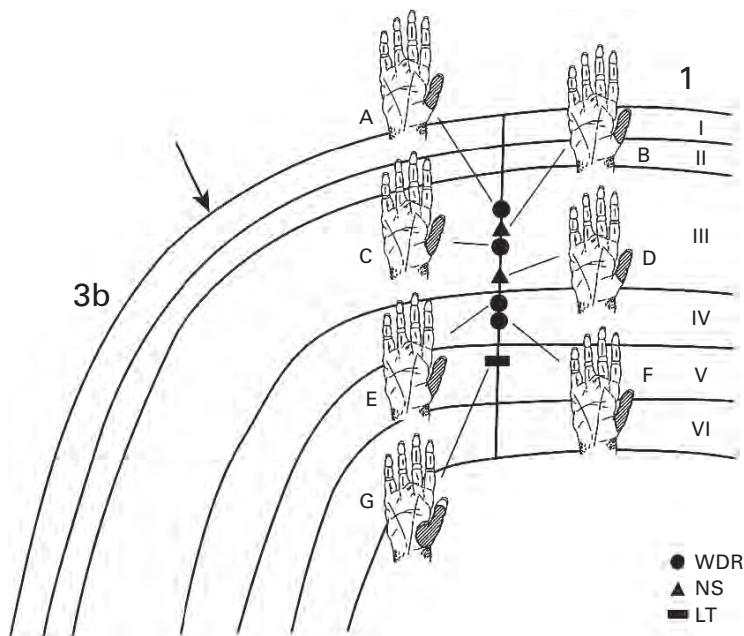
areas such as prefrontal cortex and temporoparietal regions over lower sensory regions (Desimone & Duncan, 1995; Corbetta & Shulman, 2002; Buschman & Miller, 2007; Siegel et al., 2008). The anatomical top-down feedback referred to here pertains to recurrent feedback across the cortex in general, taking effect as soon as any “higher” area is activated by feedforward input. For example, area V2 feeds its L2–3 activity back to V1 as soon as its superficial neurons start firing. No feedback from higher cortical areas is presumed necessary for this low-level feedback to occur.

6.4 Somatosensory Cortex

Because we can be equally conscious of different sensory modalities, the search for a neural basis for consciousness can be facilitated by studying whether all sensory modalities share a common circuitry. If neural-processing components strongly differ between modalities, we may ask which components belong to a common “core” and which ones are modality-specific. Somatosensation resembles the visual modality in many ways, but whereas vision denies us access to the back of an object and gives us a “2½-D sketch” at best (Marr, 1982), touch enables us to gain 3-D knowledge of objects. Given the many differences in methodologies having been used to study different modality-specific systems, this task is easier said than done.

Tactile information, coming in via the ventral posterior thalamic nuclei, is processed sequentially in subregions of SI (areas 3a and 3b to areas 1 and 2), which send their outputs to secondary somatosensory cortex (SII) and higher cortical regions, especially posterior parietal area 5. Is the local wiring diagram of SI similar to that of visual cortex (see figure 6.2)? And is a hierarchy also present in the somatosensory system?

For the local circuitry one might be tempted to say that “*l’histoire se répète*,” were it not for the fact that physiological research on somatosensory cortex began around the same time as that on the visual system (Powell & Mountcastle, 1959; Hubel & Wiesel, 1959). When letting an electrode gradually penetrate the primate cortex from superficial to deeper layers, Mountcastle and his coworkers recognized a columnar organization as also found in visual cortex. Now, however, the receptive fields of cells in a single column share a common location on the skin that is touched: cells will only fire if that particular patch of skin is stimulated, but not on another limb or part of one’s trunk or head. Cells in a single column, some 300–600 µm wide, often share a common (sub)modality. For instance, Merkel’s disks, which inhabit our subcutaneous space, slowly adapt to changes in skin pressure, channel their output into different cortical columns than outputs from Meissner’s corpuscles, which are more sensitive to gentle pressure changes such as stroking, and rapidly adapt to these. This segregation may even apply when different types of receptor are located in the same patch of skin. However, for other submodalities, such as pain, a strictly segregated columnar processing has not been found (see figure 6.4). The overall arrangement is somewhat similar to the visual cortex in that, also here, information

**Figure 6.4**

Single-unit responses to noxious stimuli were recorded while an electrode penetrated vertically through the layers of area 1 of the somatosensory cortex of an anesthetized macaque. The cytoarchitectonic border between area 3b and 1 is indicated by the arrow toward the left. Cells that are subsequently encountered in layers 3–5 share a common receptive field on the right thumb. Diverging from a classical columnar organization, different nociceptive and mechanoreceptive cells are found in this vertically oriented aggregate of cells. WDR neurons (filled circles) respond to a wide-dynamic range of stimuli, such as a brush, pressure, pinch, or thermal noxious stimulus. NS neurons are nociceptive specific, only responsive to noxious stimuli. LT neurons are low-threshold mechanoreceptive cells, responding to light touch and hair movement and not increasing their firing rate when stimulus intensity enters the noxious range. From Kenshalo et al. (2000), with permission from the American Physiological Society.

processing of some features (e.g., orientation) is harnessed into a columnar organization, whereas color information is processed via interspersed anatomical microregions called “blobs.”

The description of how somatosensory information flows within a column shows close parallels to schemes devised for the visual system. In the primary sensory cortex SI we encounter the usual suspects: thalamic afferents terminating in L4; stellate cells contacting L2–3 cells which emit projections to nearby and more distant cortical areas; and L5 cells receiving these corticocortical inputs and projecting to cortical and subcortical structures, as do L6 cells. Paralleling the recent updates in visual cortical circuitry (see figure 6.2), recent physiological studies on the somatosensory cortex of rodents—with their exquisite whisker system—have extended our picture of local connectivity (Lefort et al., 2009).

Given the vast differences between the visual and tactile modalities, and their diverging demands in processing, the global similarity between visual and somatosensory cortex is significant, even if some specific laminar connections differ. Exemplifying a deviation from canonical circuitry, Petersen and colleagues used *in vitro* patch-clamp recordings from multiple neurons in slices of mouse barrel cortex, a technique that can elucidate the strength of cell-to-cell connections with great precision. They identified strong projections from L4 cells to L5 and L6, so that L4 may not only “preprocess” thalamic information destined for corticocortical exchange via the superficial layers but also gate access of thalamic input directly to deep-layer control over subcortical targets. Using optogenetics, Petreanu et al. (2009) confirmed such patterns of connectivity, emphasizing the central role of somatosensory L3 and L5 cells in processing thalamic and corticocortical inputs.

If we zoom out from the microcircuitry to the mesoscopic scale of cortical areas, can we discern a hierarchy between somatosensory areas as deduced for vision? Resembling primary visual cortex, the “lowest” areas 3a and 3b of the primate brain contain cells with small receptive fields and very specific responses, determined by inputs often from only one type of mechanoreceptor. These cells are thought to analyze and report the fine, local details of skin contacts, whereas cells in the higher areas 1 and 2 exhibit larger receptive fields and process more complex information—for instance, on the direction of an object moving across the skin or its curvature and size. Area 1 and 2 cells integrate the finer, local information fed to them from lower areas. Hsiao (2008) conveniently phrases this difference as that area 1–2 cells have a wider “view” on the skin contacting an object. Several local edges detected across several phalanges need to be combined to sense the overall shape of a knife lying in your hand, while its sharp, serrated edge involves inputs from yet different columns in areas 3a and 3b. Up to this point, integration has been restricted to tactile inputs. But in SII and area 5, cutaneous inputs are integrated with proprioceptive inputs from muscles and tendons, offering additional information to estimate weight, size, and 3-D shape of objects (Hsiao, 2008). In posterior parietal cortex (area 7), touch information is combined with visual information to align both modalities, which is essential for the motor system to direct grasping and dexterous manipulation (Jeannerod et al., 1995; Macaluso & Maravita, 2010).

These functional indications for a somatosensory hierarchy are in line with anatomical evidence. Corticocortical tracer studies by Friedman (1983) showed that SI heavily projects to L3–4 of SII, resembling a forward projection, whereas SII projects back most densely to layer I of SI, suggesting a backward projection. By comparing laminar patterns of histochemical staining for enzymes involved in neurotransmission and metabolism, Eskenasy and Clarke (2000) found that areas 3a and 3b resembled primary visual areas, whereas areas 1 and 2 were more similar to higher visual areas. The ascending and descending relationships of areas 1 and 2 with the parietal areas 5 and 7, prefrontal cortex, and temporal lobe remain to be charted in more detail (cf. Avendano et al., 1988; Rozzi et al., 2006).

6.5 Auditory Cortex

A division between primary, secondary, and even higher areas can also be applied to the auditory cortical system. Here, primary auditory areas are distinguished from higher, so-called “belt” and “parabelt” areas in that they receive a direct thalamic input from the medial geniculate nucleus (Hackett, 2011). Whether a hierarchy exists among auditory cortical areas across mammalian species is difficult to say at present because lamina-specific tracing studies of corticocortical connections are still scarce. Nonetheless, a hierarchical scheme of 14 areas has been proposed for the cat auditory system, with the primary auditory cortex (A1) and anterior auditory field at the bottom and parahippocampal regions at the top (Lee & Winer, 2011). Also functionally, a pattern can be distinguished marked by relatively “simple” auditory features coded in A1, ascending to higher-level features in secondary auditory cortex (A2) and surrounding “belt areas,” which also incorporate inputs from other modalities such as vision (Imig & Reale, 1980; Tian et al., 2001; Scannell & Young, 1993; Wessinger et al., 2001; Rauschecker & Scott, 2009; Smith et al., 2010). As compared to A1 neurons, cells in the auditory belt and parabelt areas have longer response latencies and show wider spectral and temporal integration properties, taking into account a broader ranges of tones and arrival times (Hackett, 2011; Rauschecker, 1998). In rhesus monkeys, the lateral belt shows areal specializations identifying distinct types of social call versus their spatial localization (Tian et al., 2001). In humans, this integrative hierarchy can be extrapolated to Heschl’s gyrus, the planum temporale, and even to some extent to Wernicke’s area, which covers a large expanse of the left association cortex, close to lower auditory areas. Here, neural mechanisms analyze high-level features of perceived and articulated speech, including features such as syllables, words, prosody, emotional tone, and conceptual meaning. Widespread damage to and around Wernicke’s area results in deficits in comprehension of speech by others but also in aspects of speech production such as selection of words and phonemes. Although more anatomical and electrophysiological work is needed to elucidate a more precise hierarchy in auditory cortex, the basic layout appears to be present.

Is the micro-organization of auditory cortex also comparable to that of the visual and somatosensory system? A columnar structure has long been recognized, with inputs from the medial geniculate body of the thalamus ending in the middle layers of A1 (Jones, 2002; Linden & Schreiner, 2003). Known intracolumnar connections are globally consistent with the classic flow of information from the middle to superficial layers and from superficial to deep layers. Also the long-range connections originating from L2–3 pyramidal cells, and the subcortical projections from L5–6 cells, have been identified (Linden & Schreiner, 2003; Huang & Winer, 2000; Barbour & Callaway, 2008).

On closer inspection, however, the architecture of auditory cortex raises doubt whether all sensory areas of neocortex share a common connectional organization. For example, the middle layers of auditory cortex, in both cat and rat, lack the stellate cells that are so

distinct for the visual and somatosensory cortex (Smith & Populin, 2001). Instead, thalamic inputs mainly reach star pyramidal cells and other types of pyramidal cell in L3 and L4. Moreover, new physiological studies have mapped strong reciprocal projections from L2–3 cells to L4 and other laminae, both locally and at more remote locations (Barbour & Callaway, 2008). Along the route from the periphery to central visual and somatosensory systems, integration of inputs from the two (visual or somatotopic) hemifields is first achieved only at the level of the cortex and not subcortically, whereas auditory inputs from the two ears already converge in subcortical nuclei. Using this early binaural integration, animals can localize auditory sources based, among others, on interaural differences in response latencies to sound waves reaching the ears, and earlier detection enables faster responses such as head orientation.

Altogether, the auditory cortex deviates on several accounts from the circuits typically reported for visual and somatosensory cortex. These deviations point, at least, to “variations on a theme” in intracortical wiring that may thus not be essential for conscious experience. It should be added that conscious experiencing of, for example, sound pitch has been associated especially with higher auditory areas, including Heschl’s gyrus, the middle temporal gyrus, and superior frontal gyri (Brancucci et al., 2011). Deviations in local circuitry likely contribute to specialized analyses the auditory cortex carries out on time-varying sound spectra received from the two ears. In A1 and higher areas, these computations will result in cell characteristics that are unique to the auditory system, as they have no equivalent in other modalities (e.g., consonance and dissonance; Tramo et al., 2001). Strongly related to audition is the echo-locating system of bats harboring feature detectors for analyzing the speed, location, and moving direction of insects in midair (e.g., Tang & Suga, 2009), underscoring an even more unique specialization of an audiospatial system with features humans are very unlikely to experience directly.

6.6 Olfaction: The Oddball of Sensory Modalities

Compared to auditory cortex, brain systems for olfactory perception present us with even stronger deviations from a classic organization. Information processed by the olfactory bulb is first passed on to basal forebrain structures including the anterior olfactory nucleus, olfactory tubercle, and piriform cortex (see figure 6.5). For conscious smell, the direct or indirect projections from the olfactory tubercle and piriform cortex to the frontal cortex—especially the orbitofrontal cortex—appear most important (Gottfried, 2010).

There are only a few areas of brain research that underwent such a strong transformation in knowledge as olfactory research over the past decades. Originally, seven basic types of odor receptors were envisioned (e.g., Amoore, 1964), and specific combinations of receptor activity patterns were proposed to account for the large diversity in odors we can discriminate. Currently, no less than about 1,000 molecularly distinct odorant receptors have been identified in rodents (380 in humans), originating from a large family of genes (Buck

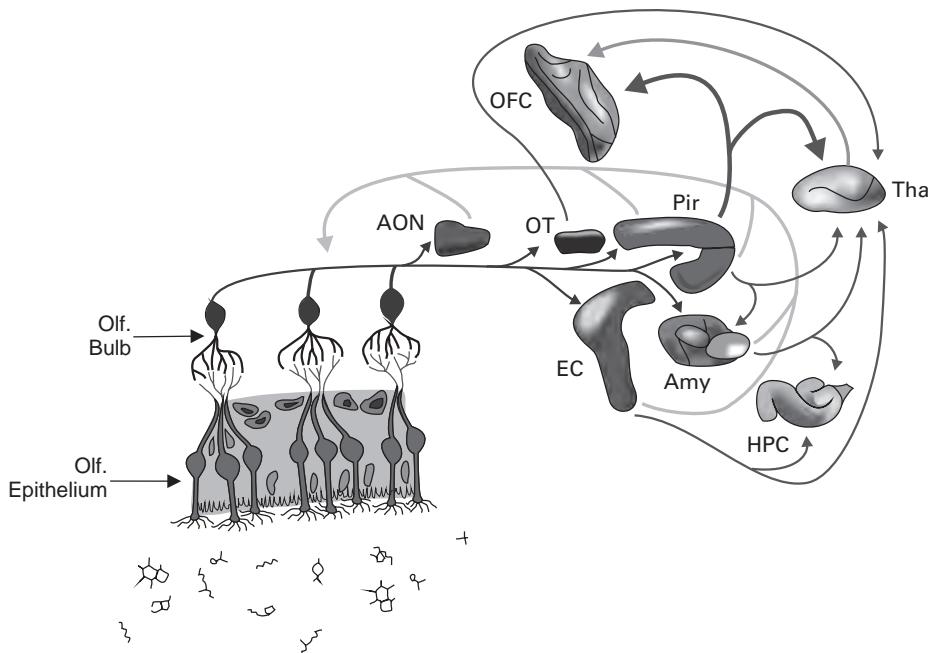
& Axel, 1991; Malnic et al., 1999). In the olfactory bulb, axons from thousands of sensory receptor cells converge on a few tens of relay cells, organized in anatomically discrete units called “glomeruli.” Neurons in these glomeruli project to the piriform cortex (Franks & Isaacson, 2006). Both across the bulb and piriform cortex, a spatially varied pattern of activity arises, a topographic “signature” of a specific odor (Xu et al., 2000; Kauer, 2002; Isaacson, 2010; but see Choi et al., 2011). Whereas the bulb may be dedicated to low-level processes such as feature extraction and selection of information, the piriform cortex may mediate higher olfactory functions involving associative memory, pattern separation, and the synthesis and distinction of odor “objects” against a background (Haberly, 2001; Wilson & Stevenson, 2003; Gottfried, 2010).

Here we stumble on some strong deviations from canonical cortex. First, anatomists have traditionally distinguished the piriform cortex as a main component of the “paleocortex” (phylogenetically “old” cortex), as it consists of only three layers instead of the usual six found in neocortex. Secondly, the mediodorsal thalamic nucleus might act as a relay for olfactory perception (Gottfried, 2010), but Ongur and Price (2000) showed that piriform–prefrontal connectivity is mostly direct; axons running indirectly via the mediodorsal nucleus constitute a minority (see figure 6.5; cf. Carmichael et al., 1994). Although this matter is not yet settled, it seems reasonable to propose that direct piriform–prefrontal connections, and not the thalamus, play a central role in conscious smell perception (Shepherd, 2007).

These observations strengthen the conjecture that a thalamic relay may be necessary for conscious vision, audition, and touch (cf. Schmid et al., 2010), but not for consciousness in general. In defense of a “canonical law” one may argue that a precortical *relay station* may be required in every modality—substituting, for example, the olfactory bulb for the thalamus—but this would belie the vast structural and physiological differences between the bulb and thalamic relay nuclei. What makes the olfactory dissidence complete, is that the caudal, olfaction-associated orbitofrontal cortex of macaque monkeys—associated with conscious smell perception with reasonable certainty (Li et al., 2010)—is *agranular*, thus lacking a layer 4 (Carmichael et al., 1994). The agranularity probably relates to the lack of primary thalamic input, and the overall data indicate that L4 stellate cells are not a necessary prerequisite for conscious perception in general.

A clear ascending–descending hierarchy has not been recognized in the olfactory neocortex. The macaque piriform cortex emits axons mainly to layer 1 of orbitofrontal cortex, and also other “forward” projections in the olfactory system terminate in superficial layers (e.g., those from the olfactory bulb; Carmichael et al., 1994).

Prefrontal areas reciprocate this input by way of output that reaches both superficial and deep layers of piriform cortex (Chiba et al., 2001). The orbitofrontal cortex is also reciprocally connected with the amygdala, medial prefrontal cortex, and mediodorsal thalamus. This reciprocity, however, is insufficient to speak of a clear hierarchy.

**Figure 6.5**

Connectivity of the olfactory (Olf.) system in the mammalian brain. Volatile chemicals bind to a large variety of odorant receptors in the nasal epithelium, where a cascade of serial and parallel processing is initiated toward and within central systems for odor perception. Diagram is mainly based on data from rodents. AON, anterior olfactory nucleus; OT, olfactory tubercle; Pir, piriform cortex; Amy, amygdaloid complex; HPC, hippocampus; EC, entorhinal cortex; OFC, orbitofrontal cortex; Tha, thalamus. Adapted from Zelano and Sobel (2005).

Functionally, firing responses to olfactory cues have been compared by Tanabe et al. (1975) across monkey olfactory bulb, olfactory cortex, and orbitofrontal cortex. These recordings revealed a gradually increasing sharpening of odor selectivity when progressing from bulb to piriform and orbitofrontal cortex. Particularly at the orbitofrontal level, olfactory information is integrated with gustatory and other modalities, as well as reward value associated with perceived odors (Critchley & Rolls, 1996; Schoenbaum et al., 2009; Van Duuren et al., 2008, 2009). Thus, the orbitofrontal cortex codes not only olfactory features of the input but also its motivational and emotional significance.

All in all, the olfactory system teaches us an important lesson. A thalamic nucleus relaying sensory information is probably not necessary to support conscious processing per se, nor is a recipient L4 in neocortex. The olfactory system provides arguments to reject the hypothesis that all sensory cortices share fundamental principles of structure and connectivity (Rockel et al., 1980), at least if these principles are specific enough to include laminar specificity and thalamic input. We may speculate that, during vertebrate evolution, the auditory and visual systems incorporated additional computational steps related to the

high speed of dense information processing in these modalities as compared to the phylogenetically “old,” perceptually “slow” and low-resolution olfactory system. However, this should not distract us from the fact that we can be intensely conscious of smell as much as of auditory or visual input.

6.7 An Interlude on Prefrontal Cortex and Consciousness

Are other prefrontal regions besides orbitofrontal cortex (and anterior cingulate cortex, in the case of affective pain) associated with consciousness? Chapter 3 presented the prefrontal cortex as a conglomerate of areas mediating cognitive functions such as attention, planning, working memory, and control over decisions, but so far its involvement in consciousness is unclear. Crick and Koch (2003; Koch, 2004) posited that the essential nodes representing specific sensory information (such as motion in area MT) must have access to prefrontal and anterior cingulate cortices via the visual hierarchy (see figure 6.3, plate 3). In turn, these frontal planning modules would provide recurrent feedback to higher-order sensory nodes, harboring higher-order, gist-like neural correlates of consciousness. In this scheme, the (pre)frontal cortex effectively becomes another essential “node” (or collection of nodes) for conscious processing. As Crick and Koch (2003) state, “The front of the brain is “looking at” the sensory systems, most of which are at the back of the brain” (p. 120).

There are several objections to this proposal. Besides the homunculoid notion of a brain area that is “looking at” inputs arriving from elsewhere, early neuropsychiatry witnessed an unfortunate period of frontal lobotomies, roughly in between the 1930s and the 1970s, inspired by the hypothesis that disorders such as depression and schizophrenic delusions were caused by aberrant prefrontal activity and connectivity (Freeman & Watts, 1937; Valenstein, 1986). Eternalized by Ken Kesey’s novel *One Flew over the Cuckoo’s Nest* and the ensuing movie, these surgeries frequently resulted in a loss of spontaneity and emotionality, decline in planning and attention—but not loss of consciousness (Mashour et al., 2005). Admittedly, a rigorous test of the hypothesis would require complete and bilateral frontal lobe ablation, which has not been documented in patients.

The prefrontal cortex is composed of a heterogeneous group of subregions having domain-specific functions. For instance, its caudal aspects are part of Broca’s area (Fuster, 2001), involved in speech production, while a more ventromedial region is engaged during memory encoding (van Kesteren et al., 2010), and the dorsolateral prefrontal cortex mediates cognitive control and working memory for exteroceptive inputs, perceived mainly by hearing or vision (Funahashi et al., 1989; Miller et al., 1996; Groenewegen & Uylings, 2000; Fuster, 2001). Parts of the orbitomedial surface of prefrontal cortex process viscero-sensory information or regulate gut motility. But whereas an orbitofrontal lesion can result in anosmia, a dorsolateral prefrontal lesion does not similarly result in blindness or deafness. Instead, one finds deficits in visuospatial working memory, which can be localized

to a specific part of a visual hemifield, possibly a “mnemonic scotoma” (Funahashi et al., 1993).

Thus, analogous to the essential nodes in the caudal sensory and orbitofrontal cortices, one might expect a modality-specific loss of consciousness with lesions along the medial and dorsolateral aspects of the prefrontal cortex, but this is not the pattern found. Even patients suffering from massive prefrontal damage were judged to be conscious (Eslinger & Damasio, 1985; Pollen, 1999; Markowitsch & Kessler, 2000; Gläscher et al., 2012; but see DeHaene & Changeux, 2011). If prefrontal cortex were equally as important for consciousness or visual attention as, for example, right parietal cortex, prefrontal lesions in the right hemisphere would be expected to have an effect comparable to hemineglect inflicted by right parietal lesions. This prediction can be made because prefrontal–parietal pathways involved in visuospatial processing are largely unilateral (Petrides & Pandya, 1984; Thiebaut de Schotten et al., 2011), but it has not been confirmed by lesion studies (Mesulam, 1981). One has to go back as far as the midcingulate cortex to find contralateral neglect in the case of unilateral lesions (Watson et al., 1973). Furthermore, connected to prefrontal areas, but located close to the insular cortex, is the claustrum, which is still enigmatic in its functions but has been linked, somewhat indirectly, to consciousness by deep-brain stimulation studies in epileptic patients (Koubeissi et al., 2014; cf. Gabor & Peele, 1964; Crick & Koch, 2005).

Although several neuroimaging and EEG studies have linked prefrontal activity to consciousness (e.g., DeHaene & Changeux, 2011), the nature of this evidence is correlative and the activity may arise from related processes such as verbal reporting, working memory, and cognitive control. When lesion evidence is added to the electrophysiological evidence, we can conclude that the current evidence pleads against a major, indispensable role of frontal cortex in consciousness—except for the significance of orbitofrontal and anterior cingulate cortex for olfaction and affective pain.

6.8 The Cerebellum and Sensorimotor Adaptation

The cerebellum, with its beautiful, almost crystalline structure, contains roughly five times more neurons than are present in the entire neocortex, yet it does not appear to make a marked contribution to consciousness. There is no shortage of sensory input or sophisticated neural circuitry in the cerebellum, and the number of synaptic processing steps is comparable to that found in neocortical circuitry. However, even massive damage to the cerebellum or its output centers, the deep cerebellar nuclei, does not cause blindness or deafness, but rather ataxia and poor learning of fine sensorimotor actions such as playing piano.¹ Dysmetria may arise in addition, which is the systematic “overshooting” or “undershooting” of limb movements aiming to reach a goal, such as hitting a soccer ball with one’s right foot, coupled to a disrupted timing of movement. At a sensory level, the ability to detect stimuli is not affected, although discrimination between similar stimuli can

be impaired (e.g., Parsons et al., 2009). Already early in life, the cerebellum is of major importance for learning to control the strength and timing of motor actions in response to sensory inputs, such as when a kid learns to climb up the stairs by making regular steps but adjusts her step size when anticipating skipping that one step with a crack in it (cf. D'Angelo & De Zeeuw, 2009). In adulthood, the brain is less in need of a cerebellum, at least as long as sensorimotor contingencies remain “calibrated” and no systematic change occurs that should prompt motor adjustment (Rambold et al., 2002; Dean et al., 2010; Norris et al., 2011).

Macroscopically, the cerebellum is divided into three anatomical zones, defined by their lateral and anterior–posterior positions: the vestibulocerebellum, spinocerebellum, and cerebrocerebellum (see figure 6.6A). Microcircuits in the cerebellum are relatively stereotyped and dominated by the Purkinje cell, the only type of neuron providing output from the cerebellar cortex (Eccles et al., 1967). Each Purkinje cell receives input from only a single climbing fiber, originating from the inferior olfactory nucleus and wrapping around the Purkinje cell like ivy. In contrast, massive inputs are indirectly provided to the cell via thousands of mossy fibers, which originate from nuclei in the spinal cord and brain stem and activate cerebellar granule cells via excitatory synaptic contacts. These granule cells, making up the vast majority of cerebellar neurons, emit parallel fibers making excitatory synaptic contact with many Purkinje cells, in a serial manner. Mossy fibers convey a mix of multimodal information to the granule cells, such as vestibular, somatosensory, visual, and auditory inputs (Apps & Garwicz, 2005). In contrast, climbing fibers are thought to convey a type of “error signal” to the *single* Purkinje cell they innervate, meaning that spikes are propagated whenever there is a discrepancy between the *expectation* (prediction) of a certain body state and its actual *occurrence* (Dean et al., 2010; but see Llinás, 2011). Here, “body state” is taken to indicate many features, including body posture, body and limb positions, as well as movement, head, or eye rotation—and even cognitive states related to expectancy of external future events. That climbing fibers convey error signals was already proposed in the computational models of Marr (1969) and Albus (1971) and has been corroborated by experimental evidence gathered over the past 4 decades.

As in the neocortex, principal excitatory circuits are supplemented with local inhibitory mechanisms, mediated by stellate, basket, and Golgi interneurons synapsing on Purkinje or granule cells. Conjoint with excitatory inputs, interneurons regulate Purkinje cell firing patterns, which are propagated to the deep cerebellar nuclei. These nuclei are thought to function primarily in the motor domain, as they control motor-related nuclei in the ventrolateral thalamus (which in turn connects to the motor cortices), the red nucleus, and vestibular nuclei.

Sensory inputs propagated via different granule cells may converge on single Purkinje cells and make them fire whenever their joint input is strong enough. However, when a parallel fiber input coincides in time with climbing fiber activity, the strength of this synapse is reduced (LTD; see chapter 2). Conversely, LTP can be elicited in the parallel fiber synapse

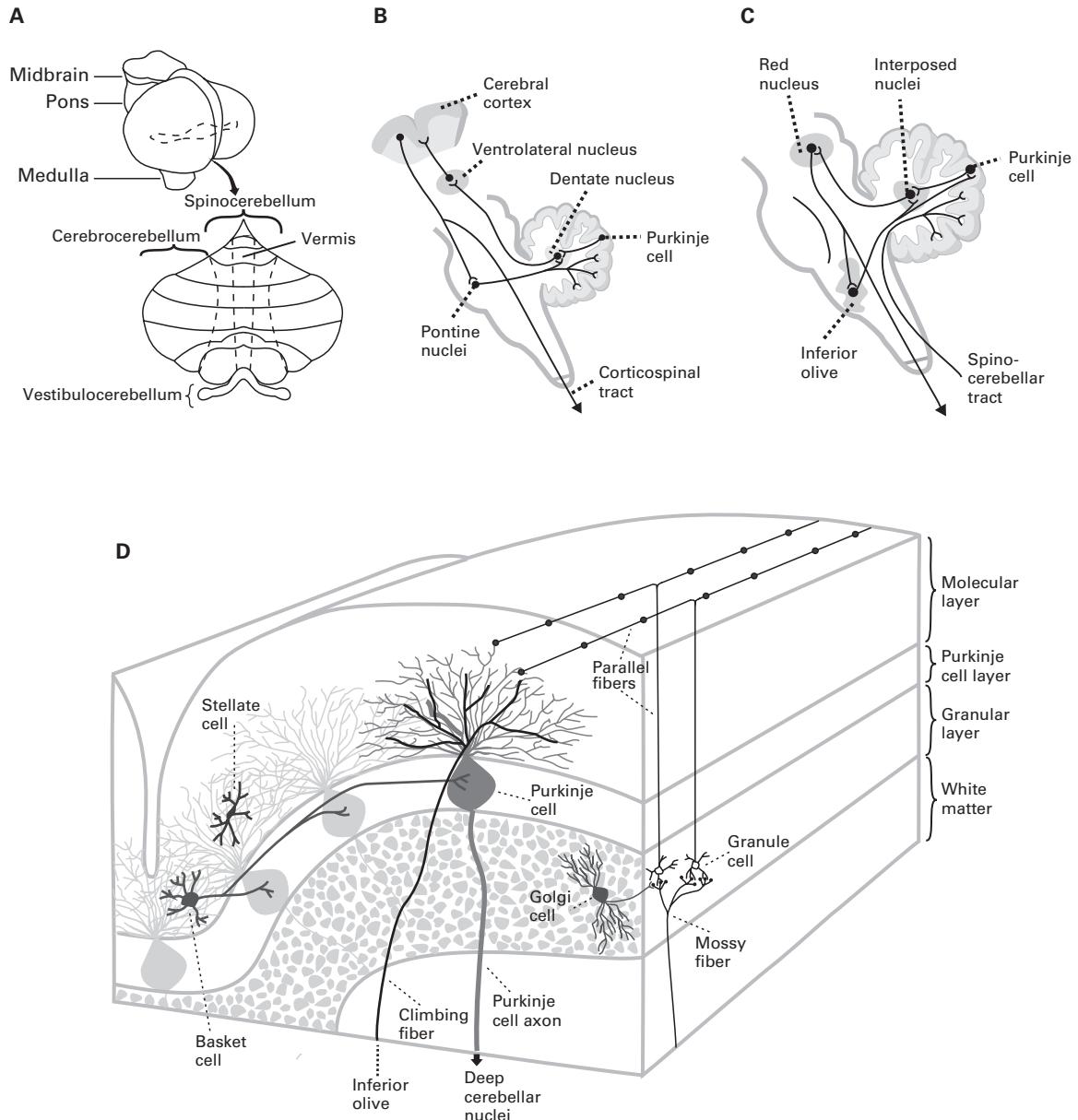


Figure 6.6

Overview of cerebellar anatomy. (A) The cerebellum is divided into a vestibulocerebellum, spinocerebellum, and cerebrocerebellum. The phylogenetically ancient vestibulocerebellum, located in a posterior strip of cerebellar tissue, functions in basic motor capacities controlling the body's balance and eye movement. The spinocerebellum lies just next to the hemispheric midline and runs from the posterior to anterior parts of each hemisphere. This zone receives somatosensory inputs via the spinal cord and controls proximal and distal muscles of the limbs and digits. The evolutionarily more recent cerebrocerebellum is located laterally in the cerebellar hemispheres and controls thalamocortical systems for voluntary movement, motor planning, and mental rehearsal of complex actions such as playing music. (B) Circuit focused on the cerebrocerebellum. Efferent corticospinal projections make contact with cells in the pontine nuclei, which feed their output into the cerebellum. Purkinje cell output reaches a deep cerebellar nucleus (dentate nucleus), which projects amongst others to the ventrolateral thalamus and indirectly to the cerebral cortex. (C) Circuit focused on the spinocerebellum—in particular, its intermediate zone. Inputs reach the cerebellum via the spinocerebellar tract; in addition the climbing fiber input from the inferior olive is shown. Purkinje cell output modulates two deep cerebellar nuclei (the interposed nuclei, consisting of the globose and emboliform nucleus), which control the red nucleus. Circuits in B and C are not complete for reasons of clarity; elements of B are also present in C (e.g., thalamocortical circuits), and the inferior olive projections shown in C are also present in the circuits shown in B. (D) Cerebellar microcircuits shown in a sliced layout. An individual climbing fiber contacts an individual Purkinje cell, which projects to the deep cerebellar nuclei. Granule cells and their parallel fiber outputs vastly outnumber the other cell types and connections in the cerebellum. (A–D) adapted from Shepherd (1983) and Apps and Garwickz (2005).

when climbing fiber activity is absent, leading to a reversal of a climbing-fiber-induced LTD effect when this was induced beforehand (Lev-Ram et al., 2003; Coesmans et al., 2004).

These and other pieces of evidence have converged to support an “adaptive filter” model of the cerebellum, which stays close to the Marr–Albus models of error-based learning. The concept of an adaptive filter can be illustrated by the vestibulo–ocular reflex (VOR), which serves the visual fixation of objects in your environment while you turn your head. Without eye movement, head rotation would cause you to lose track of an object you are watching, and the cerebellum’s job is to adjust eye movement in the direction opposite to head turns, precisely compensating for the head rotation so that the object image remains projected on a stationary portion of the retina. A “filter” is an engineering concept describing any transformation from an input (for the VOR: head velocity) into an output (compensatory eye movement). With head movement varying in three dimensions, the input is a dynamic, time-varying signal, and the filter should cope with its complex nature. The filter needs to be “adaptive,” as the parameters governing the input–output transformation may need adjustment over time—for instance, because the size and other properties of the muscular apparatus change during childhood or aging.

Here is how the cerebellum may work as an adaptive filter for the VOR. Head rotation signals are relayed as a vestibular input (V_{head} in figure 6.7C) into the cerebellum C. Conjointly, cerebellar output reaches brain stem nuclei (B) that control eye movement speed (V_{eye}). If there is a discrepancy between head and eye movement velocity, a loss of invariant image projection onto the retina (retinal slip) occurs, which is detected via the visual system and relayed to the cerebellum via climbing fibers (error signal, $error(t)$). Thus, the

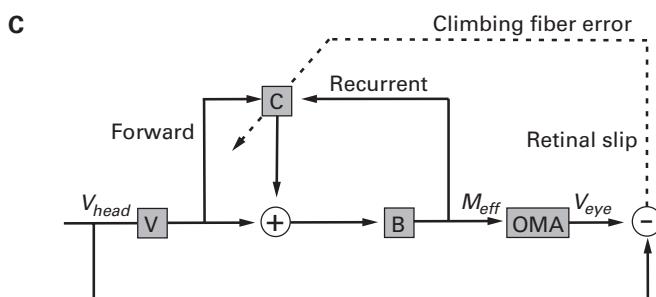
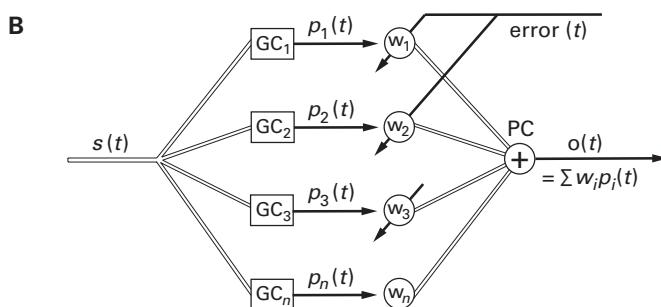
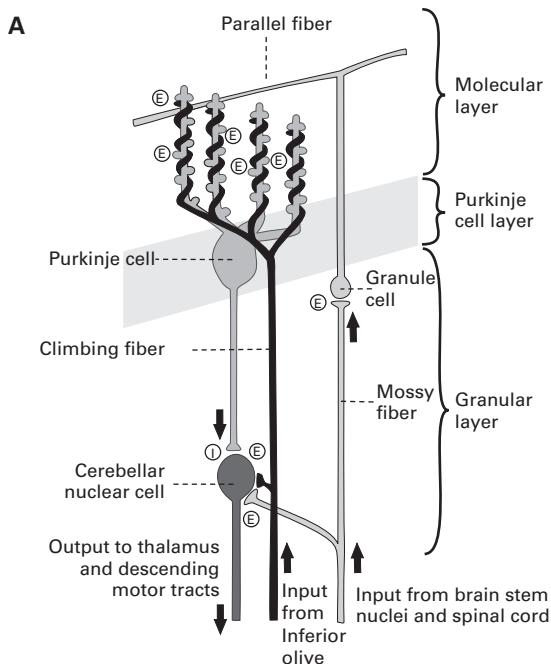


Figure 6.7

Adaptive filter model of cerebellar microcircuit. (A) Schematic version of a canonical microcircuit in which mossy fibers contact granule cells, in turn exciting (E) Purkinje cells via parallel fibers. In addition, climbing fibers emit excitatory projections wrapping around Purkinje cell dendrites. Purkinje cells inhibit (I) neurons in the deep cerebellar nuclei and thus exert indirect control over thalamocortical circuits and brain stem structures such as the red nucleus (see figure 6.6). The scheme presents a simplification in that mossy fibers inputs are thought to be distributed over many granule cells. (B) Adaptive filter hypothesis of cerebellar function. Starting on the left, mossy fiber signals, conveying time-varying sensory or motor inputs ($s(t)$), are distributed over many granule cells ($GC_1 \dots GC_n$) which project their parallel fiber outputs ($p_1 \dots p_n$) onto a Purkinje cell. The granule cells form an array of filters, each processing component signals of the original input. The strength of each parallel fiber–Purkinje cell synapse is symbolized by the weights, $w_1 \dots w_n$. These weighted inputs are summated at the level of the Purkinje cell soma (PC; + sign), with its output $o(t)$ representing the sum of weighted parallel fiber inputs. The filter is adaptive as the weights can be adjusted when a parallel-fiber input coincides in time with a teaching or error signal ($error(t)$). A learning rule specifies how an error in motor performance results in modification of the parallel fiber synaptic weight. (C) The hypothesis is illustrated by the vestibulo-ocular reflex, in which a head rotation needs to be compensated for by eye movement to keep an external object fixated. The goal is to train the circuit so that the eye-movement signal V_{eye} exactly equals the head movement V_{head} . This head movement signal is propagated forward via a vestibular unit V to the cerebellum C , which functions as adaptive filter. In addition to vestibular input, the cerebellum also receives a recurrent signal from the brain stem (B) motor output (M_{eff}), and a climbing-fiber error signal representing retinal slip. OMA represents the plant (oculomotor apparatus). The conjunction of the error signal and forward and/or recurrent signals regulates synaptic changes in the adaptive filter C . The recurrent feedback carrying an efference copy M_{eff} is thought to enable adaptations for changes in the oculomotor apparatus or brain stem circuits. Adapted from Apps and Garwicz (2005; A) and Dean et al. (2010; B and C).

forward pathway transmitting vestibular input via the mossy and parallel fibers converges onto cerebellar Purkinje cells with visual error signals.

Implementing the adaptive nature of the filter, parallel fiber synapses onto Purkinje cells can be adjusted in strength, depending on the presence of an error signal (figure 6.7B). If retinal slip is large, parallel fiber synapses will be downgraded by LTD induction so that the output of the connected Purkinje cells becomes weaker—which is functional because these cells were doing a poor job in executing the VOR. Other parallel fiber–Purkinje cell combinations do a better job when their activity does not correlate with errors. As parallel fiber activity in the absence of climbing fiber input results in LTP, this input will be reinforced.

One may wonder why such a huge number of granule cells and parallel fibers would be needed, given that a V_{head} input can be represented as a time-varying series of spikes along one fiber. A first explanation holds that the final motor command, shaped by cerebellar output, has a complex dynamic structure harboring many time constants (in figure 6.7B these are implied in the parallel-fiber output signals, $p_1 \dots p_n$). Secondly, in real life, a V_{head} signal does not come alone, because head rotation occurs in 3-D. Motor commands must take into account a vast number of input combinations, spanning multiple modalities. Also note that figure 6.7C displays a recurrent signal from brain stem motor output to the cerebellum next to the forward input, which may bring adaptive advantages, not discussed here in detail (Haith & Vijayakumar, 2009).

In summary, cerebellar learning results in control over movement not only by reflexes but also in an anticipatory, feedforward fashion, such as when we adjust our step size when

approaching an escalator. In line with this anticipatory function, it has been argued that the cerebellum computes the sensory consequences of motor actions (Izawa et al., 2012); these predictive signals may then be sent to motor centers in the brain stem to prevent unwanted outcomes. Importantly, model-building and predictive capacities appear to be at least as important for understanding the cerebellum as for the neocortex, arguing that such capacities *per se* are not key to understanding the neural basis of consciousness.

So why do cerebellar—as opposed to thalamocortical—circuits not notably contribute to consciousness? Let us first consider this from a functional, teleological viewpoint. Instead of contributing to consciousness, the cerebellum appears fit to prevent *overloading* it. For instance, patients with lesions in the lateral sector of the right cerebellar hemisphere stated that left-arm movements were executed subconsciously (i.e., automatically), but they had to think about every movement of the right arm (Holmes, 1939). The question can also be addressed from a neural network perspective. Even if the adaptive-filter hypothesis would prove false, the cerebellar system retains the basic architecture of a feedforward network, supplemented with recurrent feedback from brain stem nuclei (see figure 6.7C). In chapters 4 and 5 it was argued that such classical neural networks leave essential aspects on consciousness unanswered, so there is no discrepancy between how we think the cerebellum functions and the evidence for its role in nonconscious processing.

But then what makes the neocortex so special that it *does* play a central role in consciousness? Is there, perhaps, a crucial difference in that neocortex but not cerebellum contains sensory maps of all sorts—retinotopic, craniotopic (head centered), somatotopic? The answer is negative, because the spinocerebellum does in fact contain two somatotopic maps (Shambes et al., 1978). Moreover, having a “map” or representation *per se* is not sufficient for generating consciousness, although it may be a contributing factor. Another important neocortical–cerebellar difference might be that the cerebellum is organized in parallel, segregated circuits, in contrast to the dense interconnectivity characteristic of corticocortical wiring (cf. Tononi, 2004). However, modalities and submodalities probably affect each other, at least via cerebellar and pontine nuclei (Campolattaro et al., 2011; Apps & Garwicz, 2005). In sum, a view of the cerebellum as an adaptive filter network is not at odds with its nonconscious processing functions, but it is not yet clear how corticothalamic systems go beyond these functions.

6.9 The Basal Ganglia: The Sound of Silence

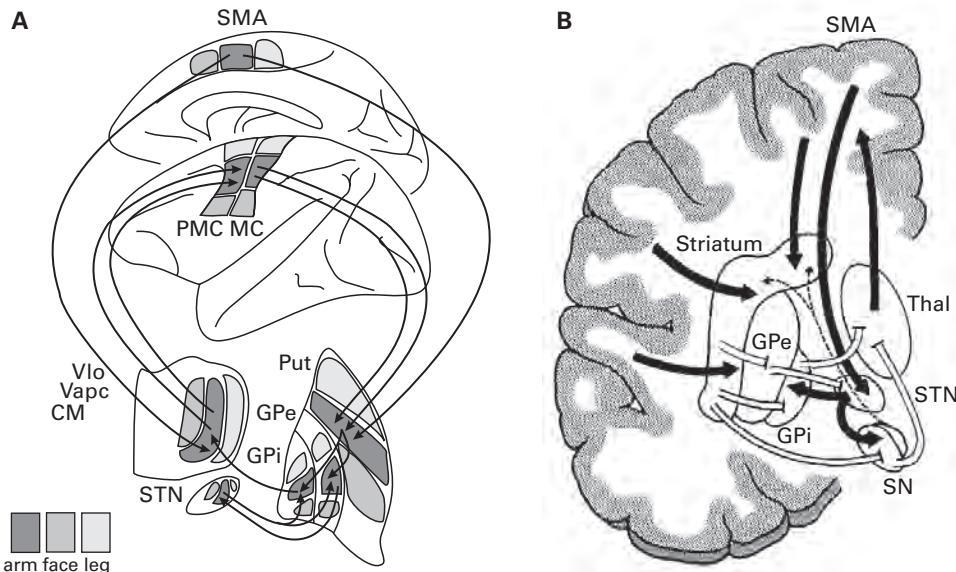
The question on nonconscious as opposed to conscious processing applies equally to the basal ganglia. Like the cerebellum, the basal ganglia have no shortage of sensory inputs, which reach this group of subcortical nuclei primarily via the neocortex and thalamic nuclei. Given that neocortical outputs—the “end results” of columnar computations, if you will—are sent to the basal ganglia via L5 cells, we face the question of how it is that

the neocortex, *generating* these end results, *is* associated with consciousness, whereas the striatum, *receiving* these end results, *is generally not*.

A key concept for understanding the organization of the basal ganglia—striatum, pallidum, substantia nigra (including pars reticulata and pars compacta) and subthalamic nucleus—is their embedding in larger anatomical constructs: loops interconnecting the cerebral cortex, basal ganglia, and thalamus. Figure 6.8A illustrates that information flow within these loops is largely unidirectional (from neocortex to striatum, but not backward, etc.) and how they are more or less “closed,” meaning that the loop’s point of origin in the cortex roughly corresponds to the locus where the loop’s thalamic output terminates (Alexander et al., 1990). Starting in the neocortex, L5 cells emit excitatory projections to neurons in the striatum, which is subdivided in the putamen, caudate nucleus, and ventral striatum. This complex contains a mixture of different cell types, but its projection neurons (see figure 2.2B) make up ~95% neurons of all cells and dominate the way information is channeled through the basal ganglia. These projection neurons inhibit their targets via GABAergic synapses.

The striatum as a whole is essentially innervated by the entire neocortical mantle, and these projections are topographically organized, such that the caudal, sensory parts of neocortex and motor cortices supply inputs to the dorsal and lateral striatum, and the more anterior, prefrontal cortical areas connect to more medial and ventral striatal sectors (Kemp & Powell, 1970; Wise & Jones, 1977; Yeterian & Van Hoesen, 1978; Nauta, 1979; Alexander et al., 1990; Voorn et al., 2004). Several loops are distinguished. For instance, two loops involving the dorsal and lateral striatum regulate movements of the eyes and skeletal musculature and originate in the frontal eye fields, primary motor cortex, premotor cortex, and supplementary motor area (see figure 6.8). This striatal sector also receives information from the posterior parietal cortex, somatosensory cortex, and occipital and temporal areas (Yeterian & Van Hoesen, 1978; Alexander et al., 1990). A different, “cognitive-executive” loop comprises the dorsolateral prefrontal cortex and caudate nucleus and mediates functions in spatial working memory, planning, and decision making (Funahashi et al., 1989; Miller et al., 1996; Fuster, 2001). Yet another, affective-motivational loop involves the ventral striatum, which receives convergent information from the medial prefrontal cortex, anterior cingulate cortex, hippocampus, and amygdala. This loop is thought to integrate information about an organism’s spatial context and environmental cues with information on reward or punishment in order to invigorate motivated behaviors (Robbins & Everitt, 1996; Pennartz et al., 1994; Pennartz, Ito, et al., 2011).

Following the loop structure from dorsal striatum to pallidum, the anatomy diverges as outputs of the caudate–putamen split up in a “direct” pathway reaching the internal segment of the globus pallidus (GPi) and an “indirect” pathway projecting to the external segment of the globus pallidus (GPe), which in turn controls the subthalamic nucleus and GPi (see figure 6.8B). The striatal-to-GPi connection is “direct” because the GPi constitutes a major output station of the basal ganglia, as it sends its inhibitory projection directly

**Figure 6.8**

Overall connectivity of the basal ganglia, with emphasis on the dorsal streams via caudate–putamen. (A) Example of a loop circuit involving motor parts of the neocortex, basal ganglia, and thalamus. Shown on top are the supplementary motor area (SMA), premotor cortex (PMC), and motor cortex (MC) which send excitatory projections to the putamen (Put). Different shading patterns indicate the various somatotopic parts of the circuits (arm, face, and leg; the diagram highlights connections between different “arm” representations). The “arm” region of the putamen projects to the external segment of the globus pallidus (GPe), but also directly to the internal segment (GPi) which inhibits thalamic nuclei that project back to the cortical areas where the loop originated. The subthalamic nucleus (STN) provides indirect connections probably involved in the pathology of Parkinson’s disease. CM, centromedian nucleus; Vapc, anterior ventral nucleus pars parvocellularis; Vlo, lateral ventral nucleus pars oralis. (B) More elaborate diagram of the connectivity of the SMA and dorsal basal ganglia, showing excitatory projections as black arrows and inhibitory projections in white. In addition to projections shown in (A), the diagram displays projections from the striatum to the substantia pars reticulata (SN), which in turn inhibits the motor thalamus (Thal). The adjacent substantia nigra pars compacta with its dopaminergic cell groups projects to the striatum (thin dashed arrows). The STN is represented as an oval region emitting excitatory projections to GPe and SN. Note that the ventral tier of the basal ganglia is not shown in this figure and is organized differently in several respects. (A) adapted from Alexander et al. (1990). (B) from Borraud et al. (2002), with permission from Elsevier.

to thalamic nuclei (Alexander et al., 1990). In addition to the GPe, the substantia nigra pars reticulata forms an output station of the basal ganglia. For the ventral striatum the distinction between a direct and indirect output pathway is less clear (Haber et al., 1990; Pennartz et al., 1994). Both ventral striatum and caudate–putamen receive strong inputs from the dopaminergic cell groups in the mesencephalon. The dopaminergic neurons of the ventral tegmentum modulate processing in the ventral striatum and prefrontal cortex, whereas the more dorsolateral neurons of the pars compacta affect dorsal striatal circuitry (Pennartz et al., 1992; Nicola et al., 2000; Hollerman et al., 2000).

Devastating neurological disorders such as Parkinson's and Huntington's disease manifest themselves most poignantly via dorsal striatal circuitry. The central causal factor in Parkinson's disease, the degeneration of midbrain dopaminergic neurons—which may be triggered by several underlying conditions—gives rise to tremors, akinesia (poor initiation of movement), bradykinesia (decline in amplitude and velocity of movement), and/or a general rigidity in posture and skeletal muscles. No loss of consciousness occurs, and thought processes remain largely unaffected, although in later stages of the disease substantial cognitive decline frequently occurs (Williams-Gray et al., 2007). Patients suffering or having suffered from *encephalitis lethargica*—“sleeping sickness”—can develop Parkinson-like symptoms in extreme form, leaving them motionless and speechless (Vilensky et al., 2010). As originally described by Von Economo in 1917 and later vividly recounted in *Awakenings* (Sacks, 1973), these patients appeared to be “mentally intact” while being inactive, sitting in rigid posture and without emotional expression on their face, illustrating how overt movement is dissociable from sustaining a mental life (Dickman, 2001).

Huntington's disease, in stark contrast, confronts a patient with an excess of movement, expressed as a “dance” (chorea) of involuntary, jerky movements of arms, legs, and facial muscles. Related hyperkinetic disorders involving the basal ganglia include ballism, manifested by uncontrolled, involuntary, even violent arm and leg movements (Comella & Shannon, 1999). On a related note, Parkinson patients can develop tardive dyskinesia as an unwanted effect of L-dopa treatment, expressed in particular by involuntary mouth and face movements outside the context of food consumption.

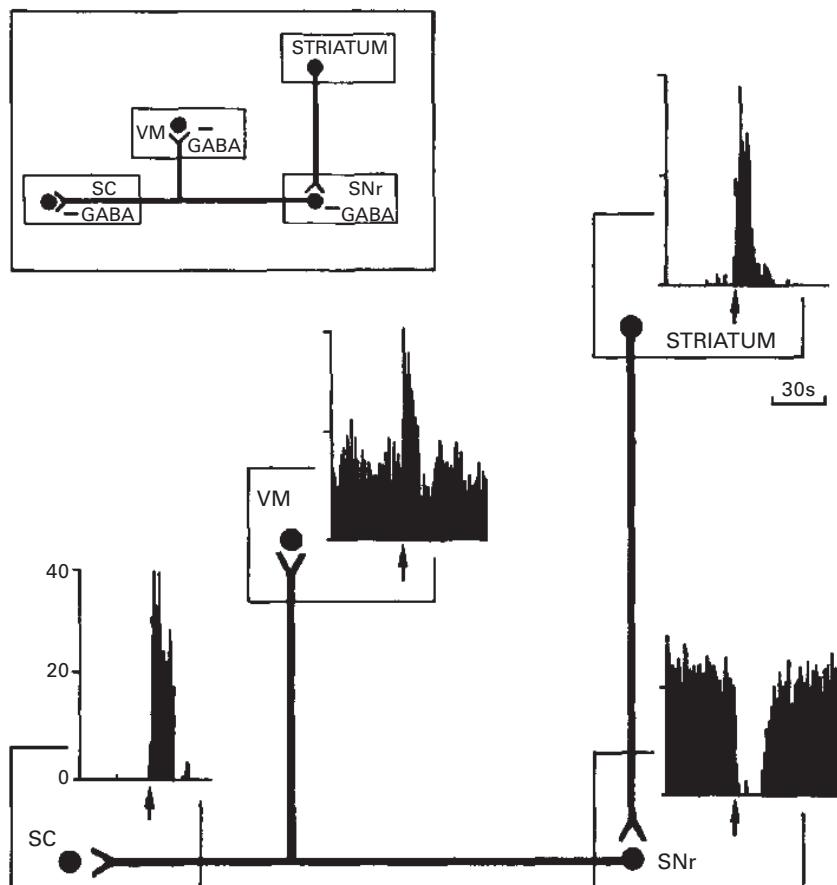
In contrast to these two neurological disorders, orbitomedial prefrontal–ventral striatal loops are implicated in neuropsychiatric pathology such as obsessive–compulsive disorders, drug addiction, schizophrenia, and depression. In humans, bilateral damage to the anterior cingulate cortex and adjacent regions can lead, among others, to akinetic mutism, expressed as a severe impairment in movement and speech (Devinsky et al., 1995; Zeman, 2001). The lack of body muscle activity is typically accompanied by a blank, vacuous expression on the patient's face, with the eyes seemingly staring into a void. Damasio (2000) argued that these patients lack consciousness (both the “core” and “extended” types of conscious states he recognizes), but it is actually not clear whether the patient's life is devoid of consciousness or of a spontaneous drive and motivation to think, move, and explore. Akinetic mutism does allow a patient to attentively track a moving object with the eyes and head and does not completely prohibit formation of memory about “vacuous” episodes, as patients can tell about their experiences later on, after their recovery (Devinsky et al., 1995; Zeman, 2001). It is thus hard to tell how akinetic mutism affects consciousness in the sense we discuss it here, emphasizing integrated sensory representation.

Returning to the dorsolateral striatum, this structure processes information received, among others, from sensory cortical areas, yet it does not appear to contribute to our conscious experience of that sensory information. In contrast, a lesion of the dorsomedial striatum has been reported to cause a hemispatial inattention or neglect deficit (Christakou

et al., 2005), raising the suggestion that a contribution to conscious processing may be made by this area. The ventromedial striatum and anterior cingulate present us with the same riddle, but now cast in subjective feelings of pain or pleasure. Anterior cingulate and related medial prefrontal activity is associated, in part, with subjective pain—the conscious sensation of full-blown, emotionally felt pain (Rainville et al., 1997; Raij et al., 2005)—whereas activity in the ventral and dorsomedial striatum is not. Ventral striatal lesions appear to lead to apathy, lack of *élan vital* and spontaneity, inattentiveness, and sometimes hallucinations and anhedonia (Mateen & Josephs, 2008). Similarly, anterior cingulate, medial prefrontal, and orbitofrontal activity has been associated with subjective feelings of pleasure and pleasantness (Berridge & Kringelbach, 2008). Indeed, patients having lesions in these areas due to defined surgical excisions report changes in feelings of happiness as well as fear, sadness, and anger (Hornak et al., 2003). In contrast, damage to their efferent targets in the striatum and pallidum fails to do so consistently, instead resulting in apathy, “psychic akinesia,” or disinhibition and impulsivity (Laplane et al., 1984; Bonelli & Cummings, 2007), in addition to motor disorders (Bhatia & Marsden, 1994). Striatal areas are strongly involved in processing and predicting reward and punishment, but the learning mechanisms to which they contribute operate nonconsciously (e.g., Pesiglione et al., 2008).

We may address this riddle from a functional viewpoint and ask what the basal ganglia’s basic job is. How are the basal ganglia designed in view of the function they are carrying out? And is this design compatible with the function of consciousness proposed earlier—the rapid construction of a multimodal representation that roughly captures what is going on in ourselves and the world?

The extensive neurological and experimental evidence on the basal ganglia boils down to *selection, regulation, and adaptation of behavior* as their primary function, supplemented with associated forms of cognitive and emotional processing. This is, frankly, the traditional notion of basal ganglia function, but I see hardly any reason to abandon it, if applied broadly enough.² The main output of the basal ganglia can be understood as *inhibiting* the flow of information exchanged between the thalamic nuclei and their cortical counterparts (see figure 6.9). GPi and pars reticulata neurons fire almost continuously at a high rate, thus shutting down their thalamocortical targets for prolonged periods of time. This tonic (near-continuous) inhibition prevents a surge of activity in the thalamocortical circuit that would otherwise trigger a movement, such as a jerk of the left arm. To prevent the system from remaining stuck in an akinetic state, an interrupt of the tonic inhibition is necessary. This is where the output from the striatum to the GPi and pars reticulata comes in. Striatal output is also inhibitory and thus suppresses the activity of GPi and pars reticulata neurons. However, for most of the time these striatal neurons are in a resting or hyperpolarized state and rarely emit spikes (Kita et al., 1984; Pennartz et al., 1994). The striatum is sufficiently quiescent to consider it the silent, dark cellar of the brain. As a result, the disinhibition effect on thalamocortical circuits only occurs at relatively rare moments when, in

**Figure 6.9**

Organization of direct basal ganglia output pathways and the disinhibitory role of the striatum. Through its “direct” pathway, the striatum sends an inhibitory projection to the substantia nigra pars reticulata (SNr), which in turn inhibits the ventromedial thalamic nucleus (VM) and superior colliculus (SC). These projections use GABA as the transmitter. When striatal projection neurons—which usually have a very low background activity—are excited by a local puff of the transmitter glutamate (time indicated by arrow), a transient firing activity is elicited. This induces a silencing of a nigral neuron, which is otherwise tonically active. As a result, the superior collicular and VM thalamic cells are released from this SNr inhibition. In this way, striatal activity can facilitate eye movement and thalamocortical processing. Histograms plot time versus spike rate (in spikes per second). Note that indirect output pathways of the basal ganglia are not shown here. From Chevalier and Deniau (1990), with permission from Elsevier.

the situation of a task, a particular motor action is appropriate. Congruent with lesioning effects, firing patterns in dorsolateral striatum code the *detailed* selection of skeletal movement, such as motion of legs, fingers, and mouth, whereas the ventral striatum codes more *global* patterns of motivated action, such as to approach or avoid a novel object in a given environment (Cho & West, 1997; Alexander et al., 1990; Lansink et al., 2012).

How the basal ganglia acquire these selecting and regulatory motor functions through learning is not yet known, but some informative observations have been made. Cortico-basal ganglia loops start out with a vast expanse of cortical inputs that converge on the striatum and thereafter are channeled toward tiny output nuclei: this architecture is evocative of an informational funnel. Hence, striatal cell groups may well engage in a competition to achieve dominance in inhibiting cell groups in basal ganglia output stations, pulling at thalamocortical strings (Groves, 1983; Pennartz et al., 1994; Plenz, 2003). A further selection may take place within these output nuclei and the subthalamic nucleus to determine which cell groups are effectively suppressed. New types of connections have been discovered that may mediate this selection function, such as inhibitory connections between striatal project neurons (Tunstall et al., 2002; Plenz, 2003; Taverna et al., 2004).

But how are the basal ganglia instructed to select an appropriate action? As touched upon in chapter 4, theorists have devised models of how the basal ganglia learn to optimize action selection and regulate the amplitude and timing of motor output. The main class of models hinges on the experimentally corroborated hypothesis that mesencephalic dopamine neurons encode an error in the prediction of future reward (Schultz et al., 1997), while other models attribute a crucial role to cortical and amygdaloid structures in reward signaling (Pennartz, 1996, 1997; Pauli et al., 2012). According to dopamine-based models of reinforcement learning, corticostriatal circuits generate *value signals* that represent the amount of reward predicted at consecutive moments in time. These signals are elicited by sensory cues or motor actions and are transmitted to dopaminergic neurons. Here, the predictive signal is compared to the *actual* outcome following the cue or action. If a discrepancy (error) between prediction and actual outcome arises, the firing rate of the dopamine neurons will change. For instance, if a reward was not predicted but yet arrived, a positive error signal is expressed as an increased firing rate. The resultant excess of dopamine release will then strengthen corticostriatal connections by promoting LTP induction, thus implementing learning. This reinforcement learning scheme for the basal ganglia shows striking similarities to cerebellar learning theory.

Having a reasonable “neuroengineering” hypothesis, we could be satisfied and state that the basal ganglia are not involved in consciousness because their job is simply not to mediate conscious representations. They subserve selection of relevant actions, and that’s it! But why could consciousness not arise *accidentally*, as a side consequence of information processing along corticostriatal loops? Revisiting the Tononi–Edelman (1998) hypothesis, it could be argued that corticostriatal loops remain largely segregated and cannot integrate the elements making up a conscious representation (cf. Tononi, 2004).

Tononi and Edelman operationalized their ideas on information integration by computing the statistical dependence between groups of neurons. If each corticostriatal loop sustains its own information flow independent of other loops, then the amount of information shared *between* loops will be low. However, a system of independently operating loops would be dysfunctional, because disinhibitory effects of one loop should be coordinated with those of other loops. Without any coordination, commands for leg movements may be disinhibited at the same time as commands for eye or mouth movements, resulting in dyskinesia or chorea.

Indeed, there is ample evidence now that corticostriatal loops interact with each other. These interactions occur at the level of striatal projection neurons, fast-spiking GABAergic interneurons, exchange between the striatum and the dopaminergic mesencephalon, or cross talk at the level of larger basal ganglia–thalamocortical loops (Haber et al., 2000; Kelly & Strick, 2004; Pennartz et al., 2009; Van Dongen et al., 2005). Overall, to ensure coordination between loops, information will be shared at multiple anatomic levels of the loop structure, accompanied by statistical dependence and complexity of information patterns in different loops.

Because striatal neurons are “silent” (i.e., not firing) in the absence of thalamocortical input, their spike patterns are highly dependent on activity in these very circuits. Thus, the cortical and striatal activity patterns within one loop will statistically be highly correlated, as confirmed by neurophysiological studies (Stern et al., 1998; Kasanetz et al., 2002; Lansink et al., 2009; Yim et al., 2011). According to a mutual-information account of consciousness, this would mean that a neocortical area and its connected striatal sector could easily form one informational compartment. Yet, neurological findings indicate that the striatal part of this compartment does not generally contribute to conscious experience, whereas in many cases the cortical counterpart does. A similar argument can be made for the dependence of sensory thalamocortical systems on inputs from subcortical sensory stations such as the inferior colliculus (for audition), trigeminal nuclei (for taste), and the retina. This indicates once more we will have to consider the representational problem of consciousness by other means than those relying on mutual information.

6.10 Sensory Processing in the Hypothalamus

Although rarely considered in studies on consciousness, the hypothalamus is no less perplexing than the cerebellum or basal ganglia. Comprising only about 0.3% of a human’s total brain weight, this diencephalic area contains a group of heterogeneous nuclei that keep our vital body and reproductive functions in balance by regulating hormonal secretion and controlling behaviors essential for survival and sexual activity. These nuclei are grouped by their position along the anterior–posterior axis of the brain (anterior, middle, and posterior regions). The anterior and middle regions guard blood and mineral supply to our internal organs by controlling blood pressure and osmolality and stimulating

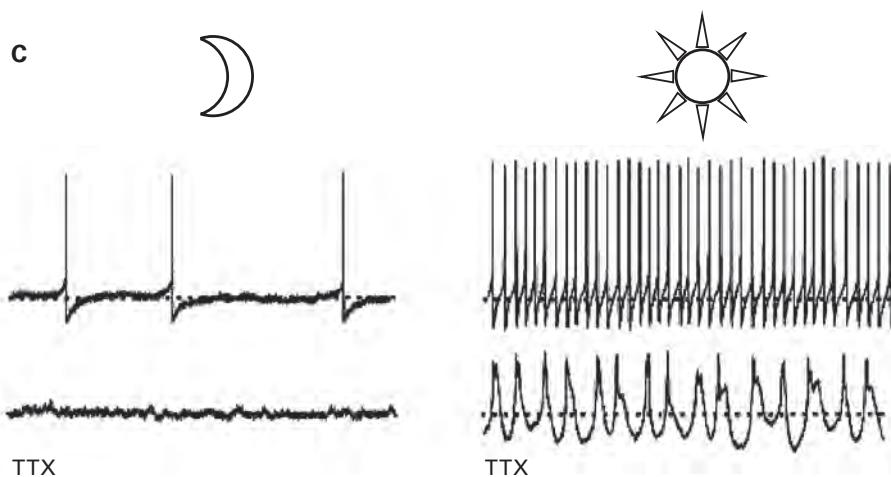
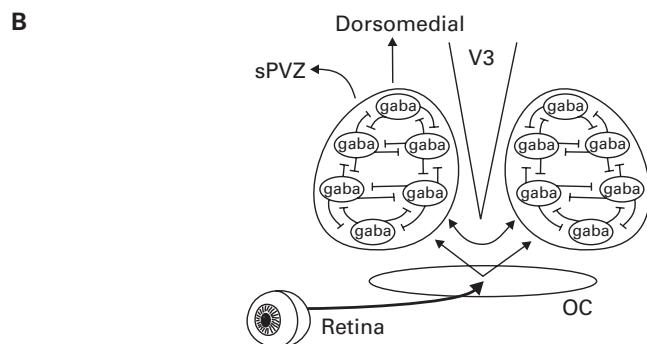
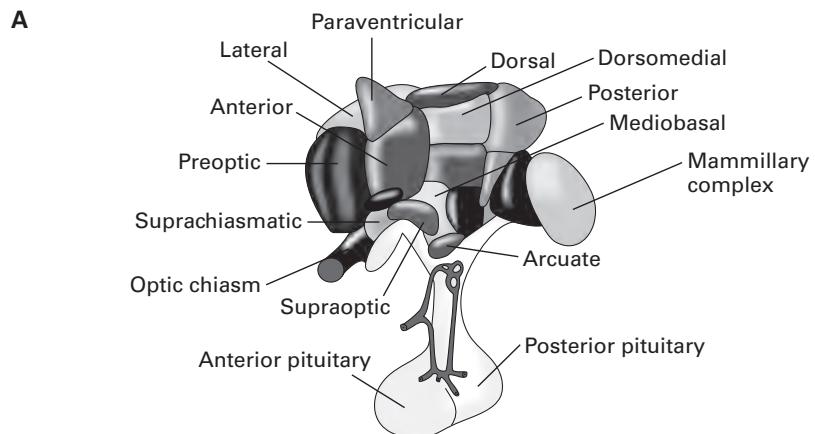
drinking or salt intake. The preoptic area of the hypothalamus (see figure 6.10) controls body temperature in addition to blood pressure. It triggers panting when it gets too hot in there, whereas other areas stimulate muscles to generate teeth chattering and shivering if the hypothalamus cools down much below a given set point (37 °C in humans; Clapham, 2012). By sensing blood glucose and other circulating nutrients, several hypothalamic areas monitor internal energy supplies and keep them in check by regulating feeding behavior, digestion, and fat metabolism. The lateral hypothalamus is famous for its triggering function in aggression (Koolhaas, 1978), and the ventromedial hypothalamus appears important for both aggression and sexual behavior (Anderson, 2012), while a host of less conspicuous nuclei in the middle region control hormones regulating pregnancy and lactation. The role of the suprachiasmatic nucleus, as our internal day–night clock, was already mentioned in chapter 2. More recently, sleep and arousal were found to be regulated by hypothalamic cells secreting hypocretin (De Lecea, 2012). The list of vital body functions involving the hypothalamus seems endless.

The hypothalamus, together with the brain stem, can be cast as a primitive “minibrain” in charge of all basic functions needed to make the organism survive and reproduce—contrasting to the “higher” cortical, thalamic, and striatal systems that blatantly execute the cognitive overhead that could be dispensed with if life itself is really at stake. When somebody loses consciousness or even sinks into a comatose state, the hypothalamus continues working with brain stem and medullar systems to control respiration, heartbeat, and blood pressure (e.g., Kinney et al., 1994).

This mesmerizing minibrain is literally stuffed with sensors and actuators that control overt motor behavior or “internal behavior” such as vasoconstriction, heartbeat, or gut contractions—yet without being involved in conscious experience. Why is this so? Why are we not explicitly aware of an internal signal telling us what time of day it is? As we did for the basal ganglia, we could work out the argument that the hypothalamus developed

Figure 6.10

Hypothalamic structures and the network of GABAergic cells in the suprachiasmatic nucleus. (A) Overview of the anatomy of the hypothalamus, with its main nuclei indicated with names. Shorthand notations have been used for the dorsal hypothalamic area, dorsomedial nucleus, mediobasal hypothalamus, posterior hypothalamic area, arcuate nucleus, supraoptic nucleus, suprachiasmatic nucleus, preoptic nucleus, lateral hypothalamic area, anterior hypothalamic area, and anterior pituitary. (B) Biological clock cells in the suprachiasmatic nucleus are predominantly GABAergic and inhibitory; these neurons also synthesize neuropeptide transmitters such as vasopressin (dorsal portion; top part) and vasoactive intestinal protein (ventral portion, bottom part). Excitatory input from the retina conveys photic input that entrains clock cells to the external day–night cycle. The mutual inhibitory connections indicate a global pattern of connectivity between groups, not implying that individual cell pairs are invariably bidirectionally connected. The nuclei of the two hemispheres are also connected by commissural fibers. V3, third ventricle; OC, optic chiasm; sPVZ, subparaventricular zone. (C) During the endogenous “night” phase of the clock cycle, suprachiasmatic clock neurons fire only sparsely if at all, and blocking action potentials with tetrodotoxin (TTX) produces a quiet baseline. During the “day” phase, clock cells show a high spontaneous firing rate, and TTX reveals an underlying oscillation in membrane potential that is independent of action potential activity. Calcium and potassium currents are thought to regulate the circadian modulation of spontaneous firing activity. (A) after Plant (2012). (C) adapted from Häusser et al. (2004).



in the course of evolution to gauge system parameters and react to sensory readings by reflexive, nonconscious actions—via hormones, autonomous nervous control, or our muscular contractions. The system has an undisputed evolutionary purpose, yet without being associated with consciousness. Hypothalamic cells gauging temperature are functionally or teleologically comparable with a household thermostat, also equipped with a mechanism to compare ambient room temperature with a set point we find agreeable to live with. There is no reason to believe that either system generates conscious experience unless we would resort to a form of panpsychism.

In contrast to a thermostat, we do have to credit the hypothalamus for its widespread connectivity with other brain structures. The hypothalamus is modulated by prefrontal and ventral striatal output, and the system affects regions in the brain stem, basal forebrain, amygdala, and cerebral cortex via multiple routes. That humans can lose consciousness while retaining an active hypothalamus can thus not be explained by any lack of connections with the rest of the brain. Instead of external connectivity, we may ask whether cells within a single hypothalamic nucleus are connected, or all act individually. As a representative (Van den Pol & Tsujimoto, 1985) example, the suprachiasmatic nucleus is composed of inhibitory, GABAergic neurons that influence each other's activity on a fast time scale. This may help them to stay synchronized in expressing the same electrical rhythm during the day–night cycle (high firing rate during the daytime, low at night; De Jeu & Pennartz, 2002; figure 6.10B). Other examples of a dense intrinsic hypothalamic connectivity are provided by the supraoptic area and lateral hypothalamus, and a major share of this connectivity is GABAergic (Decavel & Van den Pol, 1990; Obrietan & Van den Pol, 1995; Hart-hoorn et al., 2005; Armstrong et al., 2010; Higuchi & Okere, 2002). In sum, the statistical dependence between hypothalamic neurons will be substantial and probably comparable to corticothalamic systems, and the great variety of sensors and effectors ensures a high degree of differentiation.

So what *could* explain the firm commitment of the hypothalamus to maintenance of life, but not consciousness? The anatomy and physiology of the system provide important suggestions. Despite its dense connectivity, the adult hypothalamus is dissimilar to thalamocortical systems in that its intrinsic connections are predominantly inhibitory³ (e.g., Decavel & Van den Pol, 1990; Obrietan & Van den Pol, 1995), as illustrated by suprachiasmatic physiology. Thus, one of the anatomical leads to follow is that corticothalamic, but not hypothalamic, systems have the ability to propagate excitation in a repetitive and potentially reverberatory fashion, from subregion or ensemble A to subregion B, which subsequently activates C and next back to A, and so forth (cf. Abeles, 1991). In contrast, a surge of hypothalamic activity will likely extinguish when excited cells transmit spikes to neighboring cells or areas. Insofar as consciousness depends on propagation or reverberatory activity across multiple stations or processing layers, the functional connectivity of the hypothalamus suggests why this brain organ remains uninvolved.

A second viewpoint on the case of the hypothalamus *versus* consciousness departs from the idea of representational function. Whereas above the function of consciousness was cast as a quick, multimodal representation of one's current situation, a conspicuous feature of hypothalamic function is that the sensors operate on a *slow* time scale and are invariably *low dimensional*. For instance, cells in the paraventricular and supraoptic nuclei that secrete the antidiuretic peptide vasopressin are sensitive to the osmolality of blood (Prager-Khoutorsky & Bourque, 2010), but this slowly fluctuating parameter is captured as a scalar value (e.g., 286 mOsm/kg). A rising blood osmolality may prompt water-seeking behavior, but this can occur on a time scale of minutes, whereas seeing and dodging a rapidly approaching vehicle on a street should be fixed within a split second. High blood osmolality triggers a sense of thirst, but the conscious sensation of thirst—and its disappearance when subjects can drink water to satiety—has been related to activity in the anterior cingulate cortex (Egan et al., 2003). What we become aware of is not that our blood itself feels “dry” or “salty.” Upon drinking to satiety, the feeling of thirst disappears quickly, along with a decreased activity in the anterior cingulate, whereas the Na^+ level in the blood remains at a high level for some minutes after drinking, along with persistent activity in the medial hypothalamus. The same argument applies to blood glucose levels, which we cannot taste, the time-keeping signal of the suprachiasmatic nucleus and other low-dimensional hypothalamic body readings.

Brain systems involved in conscious experience appear to lack the capacity to incorporate these hypothalamic signals directly into a conscious representation. Hypothalamic sensing appears impenetrable to consciousness. Even stronger, we may say that a sensory representation should *not* purely consist of a low-dimensional signal on its own to be eligible for conscious experience. In daily, regular experience, the embedding or integration of the signal in a wider spatiotemporal context appears to accompany awareness of that signal. Systems for conscious representation are unable to work *without* a signal's integration in a context. As a mental exercise, try to imagine what high osmolality would feel like in the absence of *any* bodily context. So far, the importance of context has been implicit in the concept of multimodal representation of one's current situation, and the hypothalamus poignantly illustrates why the situational aspect is a key property of that function. Philosophically, the idea that experience is set within a spatial context pays tribute to Kant's (1787) *a priori* conditions of space and time.⁴ It is the hypothalamus, however, that provides a powerful neurobiological underpinning of this notion. Similar arguments can be made for the enteric nervous system (Lam, 2010), brain stem, medulla, and the connected autonomous ganglia.

6.11 What Is So Special about Corticothalamic Systems?

Having ventured into the subcortical realms of the brain, we must ask once again what is so striking about corticothalamic systems that it would underpin their privileged

association with consciousness. Both the hypothalamus and basal ganglia are marked primarily by intrinsic inhibitory circuits, but this does not generalize to the cerebellum. At the same time, however, cerebellar circuitry has a predominant feedforward structure, whereas at least the visual and somatosensory cortical hierarchies have feedforward as well as recurrent pathways. Could “recurrency” in a network be a key property distinguishing corticothalamic systems from subcortical systems? Electrophysiological studies on the visual system indicate the potential importance of recurrent feedback from higher to lower cortical areas (Lamme & Roelfsema, 2000; Lamme, 2006; DeHaene & Changeux, 2011) although there is no conclusive proof that recurrent feedback would be a key requirement for consciousness.

Here, I argue that recurrence per se is unlikely to qualify as a key property although it may turn out to be a necessary requirement. First, recurrent neural networks are under-constrained when it comes to explaining consciousness (see chapters 4 and 5). Second, both the basal ganglia and cerebellum do have a certain degree of recurrent feedback. In the basal ganglia this is apparent, for example, from some projections between individual structures, such as the pathways from striatum to pallidum and back (Kuo & Chang, 1992; Bevan et al., 2002). Moreover, overall the corticostriatal loops have a partially recurrent nature, as the thalamic output of a loop approximately reaches the area where the same loop originated. As for the hindbrain, brain stem nuclei send recurrent projections back to the cerebellar Purkinje cells (see figure 6.7C).

The unique architecture of the neocortex becomes clearer when we take a step back from the classical “vertical,” columnar view on cortical organization (see figure 6.2), and likewise from a strict feedforward-recurrent interpretation of cortical hierarchies (see figure 6.3, plate 3). Kandel et al. (2000) state the following:

The columnar organization of the cortex is a direct consequence of cortical circuitry. The pattern of intrinsic connections within the cerebral cortex is oriented vertically, perpendicular to the surface of the cortex. [...] This means that the same information is relayed up and down through the thickness of the cortex in a columnar fashion. (p. 457)

This view may be valid for *local processing* in the cortex but is less helpful in understanding *global* functions of the cortex, especially the construction of multimodal, situational representations. Here, an ordering and integration of information coded across many different regions is thought to be required, and so we must rotate our view on the neocortex by 90° and focus on the web of corticocortical connections tangential to its surface. The diagrams of figure 6.2 zoomed in on local circuitry, and thus it remained underexposed that these connections often run across a long range, comprising both interhemispheric and ipsilateral projections. These long-range projections allow almost any cortical area to affect processing in another, distant cortical area in a few steps. The overall structure of the neocortex possesses characteristics of a small-world network (Watts & Strogatz, 1998; Buzsaki, 2006; Sporns et al., 2007). The feasibility of an easy access, rapid information

exchange between cortical nodes also becomes clearer if the “lateral” connections between cortical areas (see figure 6.3, plate 3) are highlighted. Generalizing from lower sensory cortical systems to higher associational areas, where an ascending–descending hierarchy is much less clear, an alternative diagram of “horizontal” cortical organization is shown in figure 6.11.

This more generic scheme shifts our attention away from mere recurrency toward reverberatory capacities of the system. Subcortical structures lack this architecture for *very* widespread, mutual exchange between regions and virtually unlimited reverberation. Concurrent with reverberatory capacity, the system has a potential to generate recursive patterns of activity (cf. Treves, 2005). Originating from linguistics and mathematics, *recursion* is a process that occurs when a procedure which is followed to produce a certain outcome relies on invoking this very procedure itself, usually repetitively and in a self-similar way. Following Fibonacci’s example, where the n th member of the sequence is computed as the sum of the members $(n - 1)$ and $(n - 2)$, figure 6.11 suggests a scheme by which the results A and B from the respective nodes can be used to compute a result C at a time t , which in turn propagates to A , B , and other nodes to recompute their new states at $t + 1$. This capacity may come in very handy when studying how situational representations arise and evolve across time and are subject to specification and reinterpretation.

How important are corticocortical connections for building representations as compared to the sensory thalamic inputs to the neocortex and to local intracortical connections (see figure 6.2 and 6.4)? If functional importance is mirrored by the density of anatomical connections, the suggestion emerges that corticocortical connections contribute the lion’s share as they make up a larger portion than originally estimated from tracing studies. Less than 1% of the total number of cells that project to V1 lie in the LGN (Douglas & Martin, 2007). The vast majority of L4 inputs are believed to originate from L6 and other spiny stellate cells in L4, partly in the same microregion, partly in other areas. Particularly for L1 and L6, a “dark matter” problem exists in that only a small portion of synapses has been traced back to an identifiable origin that is either thalamic or local cortical (Vezoli et al., 2004; Douglas & Martin, 2007), and the contribution of remote corticocortical connections is probably larger than previously thought (Braitenberg & Schüz, 1991; Laramée et al., 2012). Thus, neuroanatomy supports Braitenberg and Schüz’s (1991) idea that the cortex primarily talks to itself. Indeed, the “horizontal” view of the cortex emphasized here advocates a busy, multidirectional information exchange between cortical areas, interrupted only occasionally by the “needle prick” of a thalamic input to L4.

The importance of long-range corticocortical connectivity was also recognized in earlier theories of consciousness (e.g., Crick & Koch, 2003; Tononi & Edelman, 1998), especially in the global workspace theory of DeHaene and Changeux (2011). The main difference between global workspace theory and the current framework is that the first gives an account primarily of how executive cognitive modules operate on sensory representations, and how representations are processed to guide behaviors, whereas here we concentrate on

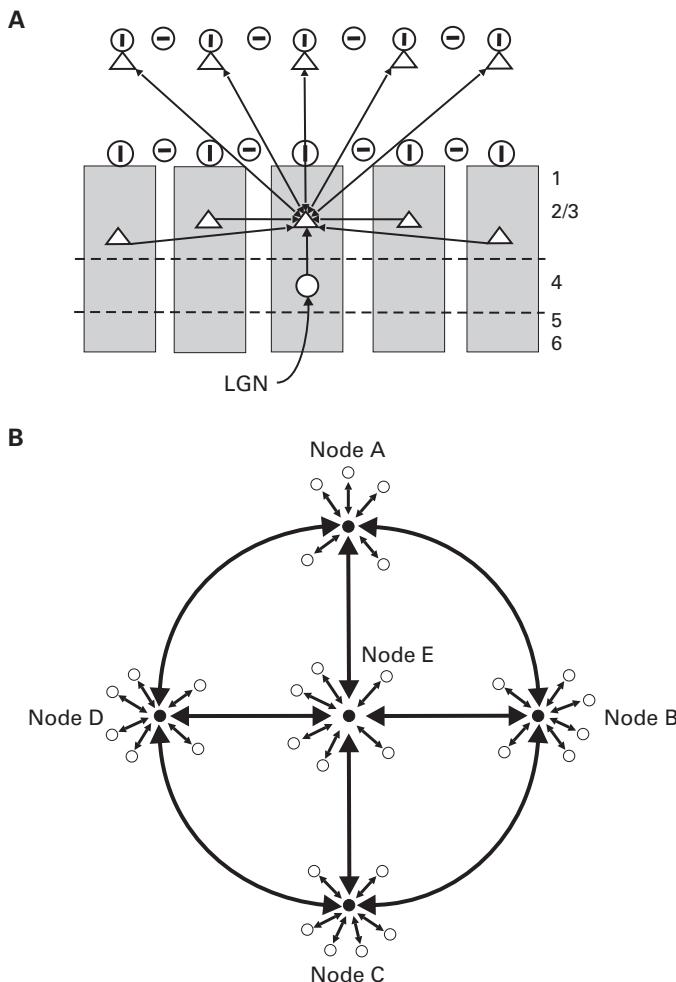


Figure 6.11

Short- and long-range horizontal communication in the cerebral cortex. (A) Scheme of local horizontal connections in macaque primary visual cortex. Light gray zones depict anatomical columns sharing a preference for vertical line orientations of stimuli appearing in the receptive fields of layer 2/3 neurons. The neuron in the centrally placed column receives horizontal inputs from similarly tuned neurons in nearby columns. Except for the main lateral geniculate nucleus (LGN) input to layer 4, no other intracolumnar circuits are shown here (see figure 6.2). Open triangles in the upper row represent neurons in the extrastriate cortex, also proposed to have similar orientation tuning, which feed back on the centrally placed neuron. These feedback connections may, amongst others, provide a surround modulation of responses to stimuli restricted to the neuron's receptive field. (B) Scheme of intra-areal and interareal cortical connectivity viewed in a tangential plane. The five nodes shown represent clusters, each of which contains similarly tuned, commonly activated neurons. These neurons communicate with other neurons in the same cluster by short-range (intra-areal) connections. Each node harbors a central neuron (filled circle) providing interareal communication. Unlike (A), this communication does not assume that only similarly tuned neurons are connected because different nodes may code for different sensory attributes (e.g., color, shape, motion). For simplicity the scheme focuses on corticocortical communication via the superficial layers, while deep-layer projections and interneurons have been left out. The resulting architecture resembles a small-world network. Examples of nodes showing this type of connectivity include, for instance, area V4, MT (middle temporal), inferotemporal cortex (IT), and the secondary auditory and somatosensory cortices. (A) adapted from Lund et al. (2003).

the construction of representations *per se*.⁵ If we pursue the matter of consciousness as a problem of multimodal, situational representation and accept that the required integration cannot be achieved by unconnected feature detectors (see figure 4.10), the question arises as to how sensory neocortical areas processing different sensory modalities are actually interconnected.

Cortical anatomy has raised evidence for two basic types of connection between sensory areas of different modality. The first type is direct connectivity between sensory areas that have traditionally been considered unimodal. For instance, connections have been demonstrated between visual and somatosensory regions, and between visual and auditory areas (Dehay et al., 1988; Schwark et al., 1992; Huffman & Krubitzer, 2001; Schroeder et al., 2001; Cappe & Barone, 2005; Falchier et al., 2002; Ghazanfar & Schroeder, 2006; Schroeder & Lakatos, 2008).

The second type of configuration for cross-modal information exchange is indirect. Several configurations are known in which two cortical areas of different modality send converging projections to a common, higher-order sensory area, the nature of which is multimodal (or “polysensory”). Examples of such multimodal convergence areas are the superior temporal polysensory area, ventral intraparietal area, and inferior parietal lobule, which includes the lateral intraparietal area (LIP; Bremmer et al., 2001; Calvert, 2001; Schroeder & Foxe, 2002; Cohen et al., 2004; Avillac et al., 2005). Here, visual information converges on groups of neurons that also receive, to varying degrees, auditory, somatosensory, and vestibular inputs (e.g., Bruce et al., 1981). Neurons in these areas have been associated with multiple integrative functions, such as prediction of sensory events based on inputs from different modalities, improvement of sensory detection based on combined modal inputs, and spatial orientation behavior.

Also in front of the central sulcus, cells in the monkey or human ventral premotor cortex are sensitive to proprioceptive, visual, and auditory inputs, particularly those informative about the spatial positions of limbs and external objects (Graziano, 1999; Bremmer et al., 2001; Lloyd et al., 2003). Further examples include the occipitotemporal and temporoparietal junction areas integrating visual and tactile information (Amedi et al., 2001; Arzy et al., 2006) and the insular cortex (see chapters 2 and 3; Calvert et al. 2001). Several long-range, cross-modal connections terminate in deeper (infragranular) cortical layers, suggesting a feedback-style information flow, but exceptions have been found and there appears to be no general rule to classify multimodal connectivity as “ascending” or “descending” (Schroeder & Foxe, 2002; Falchier et al., 2002; Cappe & Barone, 2005; Ghazanfar & Schroeder, 2006).

Although the mere existence of cross-modal connections and interactions does not prove any particular view on consciousness *per se* (because they might in principle also support other, nonconscious functions), I will argue in chapters 8 and 9 that they provide one of the main cornerstones for conscious representation.

To conclude, we have seen how local circuit and area-to-area connectivity patterns across the neocortex appear similar at first glance but fail to generalize across all sensory modalities (with olfaction as the notorious oddball) and fail to explain, on their own, why this brain structure would uniquely contribute to consciousness. Also subcortical regions such as the cerebellum, basal ganglia, and hypothalamus harbor sensing and information-processing principles that have been previously considered essential for conscious representing (e.g., recurrency, statistical dependence, and complexity). To move forward on the concept of conscious experience as instantiated by multimodal, situational representation, it is important to frame-shift from a “columnar” as well as strictly hierarchical view to a more “horizontal” view that emphasizes the unique small-world, reverberatory and recursive properties of the fundamentally excitatory cortical network. We will explore later whether such properties may support construction of higher aggregate forms of representation from lower-level forms operating at the level of single neurons and within-area groups of neurons.

7

Same Brain, Different States: Waking, Sleeping, and Anesthesia

And I must reject all the doubts of these last few days, as hyperbolical and ridiculous, particularly the general uncertainty about sleep, which I could not distinguish from the wakeful state: for now I see a very notable difference between the two states in that our memory can never connect our dreams with one another with the general course of our lives, as it is in the habit of connecting the things which happen to us when we are awake.

—R. Descartes (1641, p. 168)

7.1 Different Brain States across the Sleep–Wake Cycle

Having seen that different brain systems show a distinct involvement in conscious experience versus nonconscious representation, we will next examine dynamic states in which brain systems express this involvement. Corticothalamic systems may have the appropriate long-range and local circuit connections to process sensory information consciously, but in what kind of state should they be to do so? I will review global EEG states of the brain as being differentially associated with conscious experience—wakefulness, REM sleep, non-REM sleep, and anesthesia. REM sleep is often coupled to dreaming activity, a brain state resembling wakefulness but also distinct in several ways. This makes it fascinating to find out what is different about the neurophysiology of wakefulness versus REM sleep. We will also focus here on a form of neural activity found in non-REM sleep called “replay.” This phenomenon pertains to the internal recurrence of experiential records that were acquired during preceding learning in the awake state. If such experiences acquired during wakefulness return during non-REM sleep, is this coupled to a conscious reexperiencing? Finally, we will scrutinize the fast time scale of dynamic processing in the awake state, as encountered during conscious versus nonconscious processing. Here, cortical activity states are distinguished by fast brain rhythms such as gamma activity, and we will ask which of these rhythms or processing modes may determine whether you will see a briefly flashed stimulus or will miss it.

When dozing off, your brain undergoes a profound transition that is reflected in electrical measures of eye movement, skeletal muscle activity, and EEG activity recorded from

the scalp (see figure 7.1A). In the 1950s Kleitman, Dement, and colleagues discovered that sleep is composed of two very different phases, REM and non-REM sleep (Aserinsky & Kleitman, 1953). When you fall asleep, your brain proceeds from waking into progressively deeper stages of non-REM sleep. These stages are labeled 1 for the initial, relatively “superficial” phase through 4 for the late, “deep” stage. They are not discretely segregated but gradually merge into one another as time goes by. After 1–2 hours of non-REM sleep the brain usually climbs out of this state to arrive in REM sleep (see figure 7.1B). In addition to being characterized by rapid eye movements, REM sleep can be recognized by a loss of body muscle tone, whereas during non-REM sleep skeletal muscles remain somewhat active.

Before going into detail about EEG characteristics of wakefulness and sleep, let me first explain more about how EEG traces are recorded. Scalp EEG is measured by placing one or more electrodes on the head’s skin. An EEG trace is defined as a continuous line of voltage fluctuations over time, recorded by a single electrode, and this voltage is measured with respect to an electrode placed at another point on the body surface that is insensitive to cortical activity and thus “neutral,” called the reference electrode. Beneath the measuring electrode, large groups of cortical neurons generate voltage changes in the extracellular space surrounding them. This is possible because excitatory postsynaptic activity is predominantly made up of positively charged ions (Na^+) flowing through glutamate receptor channels from the extracellular space into the neuron. This influx depolarizes the neuron but concomitantly causes a slight negative voltage deflection in the space directly outside due to the local loss of positive charge.

Spikes are made up of a very rapid flux of Na^+ ions into the cell, and also this flux is coupled to a net loss of positive charge directly around the soma and axon. If many neighboring neurons exhibit the same type of activity and have their dendrites oriented similarly, the tiny negative voltage deflections add up across extracellular space, and the field effects resulting from this population activity are measurable through the skull. However, even if one measures EEG activity from many points across the scalp, it is very difficult to pinpoint exactly where the source of an EEG change is anatomically localized. Below I will also refer to local field potentials (LFPs), which are voltage traces obtained by the same type of extracellular measurement as scalp EEG, but now the recording electrode is placed inside a brain area of interest, in the midst of neurons generating the recorded signal.

The scalp EEG of non-REM sleep is characterized by delta oscillations (1–4 Hz; see figure 7.1; often frequencies $< 1 \text{ Hz}$ are considered separately and labeled “slow oscillations” or “slow rhythm”; Steriade, Contreras et al., 1993; Steriade, Nuñez, & Amzica, 1993; Crunelli & Hughes, 2010). These rhythmic voltage fluctuations are reflections of corticothalamic neurons which go, massively and synchronously, through cycles of activation (depolarization) and inactivation (hyperpolarization), lasting about 0.5 to 2 seconds each (see figure 7.2A). Slow oscillations gradually grow in amplitude as stage 1 sleep progresses into stage 4. Hence, non-REM sleep is frequently called “slow-wave sleep,”

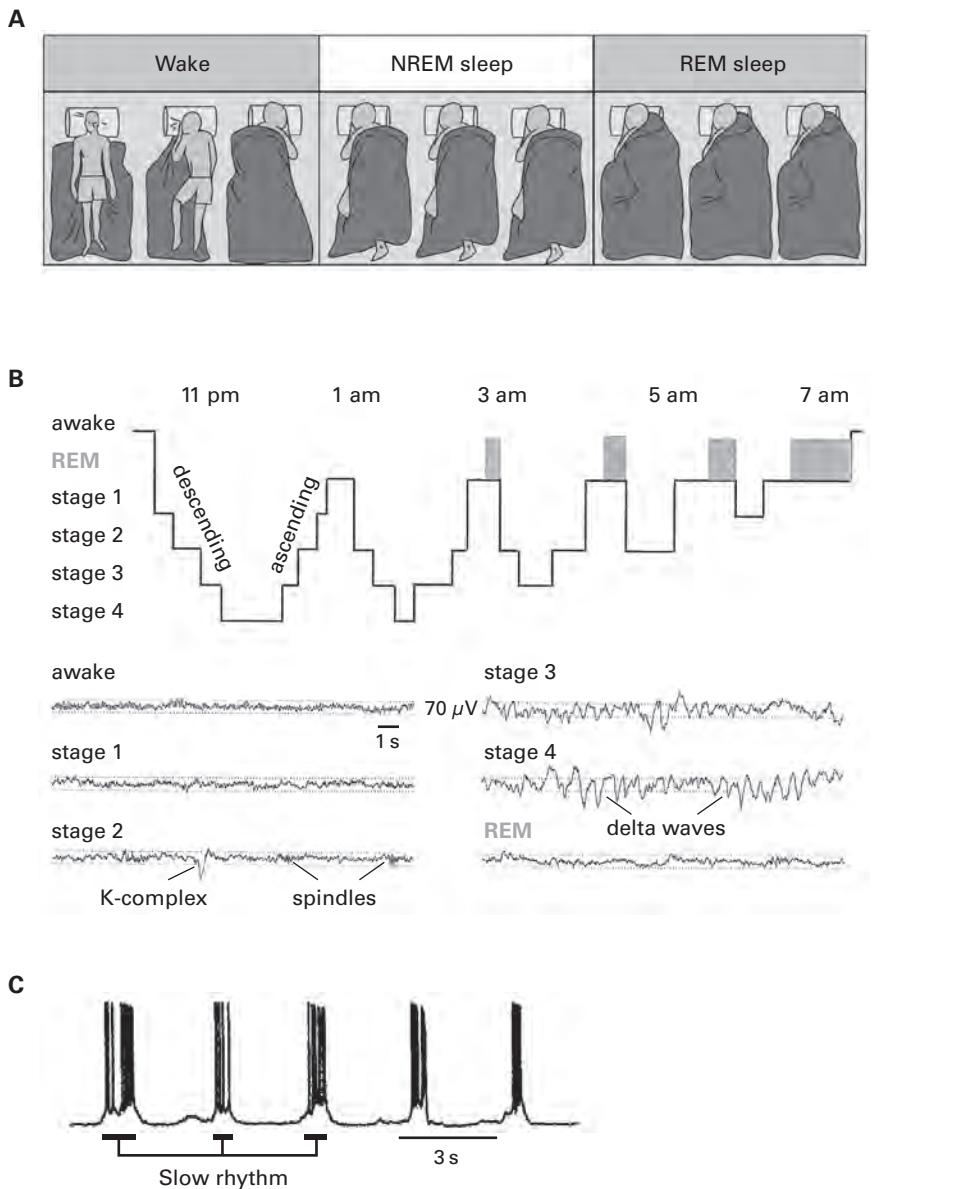


Figure 7.1

Periods of wakefulness, non-rapid eye movement (NREM) sleep, and REM sleep (A) are marked by distinct EEG patterns recorded on the scalp (B). These periods alternate across several cycles throughout the night (B, upper panel), with REM sleep becoming more prominent as the night period progresses. EEG fragments show how the desynchronized pattern characteristic of wakefulness changes into an EEG pattern displaying waves, becoming stronger and more frequent from stage 1 to stage 4 of non-REM sleep. Note spindles and K-complex occurring during stage 2 (early non-REM) sleep. The EEG pattern of REM sleep resembles the awake state. (C) In addition to the slow (delta) waves marked during stage 4 in (B), a slow (<1 Hz) rhythm in the cortical EEG has been demonstrated both during natural sleep (Steriade, Nuñez, & Amzica, 1993) and anesthesia. Here an example is shown of a slow (~ 0.3 Hz) rhythm intracellularly recorded from a pyramidal cell in cat neocortex under anesthesia. (A) from Hobson (2005). Reprinted by permission from Macmillan Publishers Ltd. (B) from Buzsaki (2006), including EEG traces from the lab of A. A. Borbély, reprinted with permission from Oxford University Press. (C) from Steriade, Contreras, et al. (1993), with permission from the *Journal of Neuroscience*.

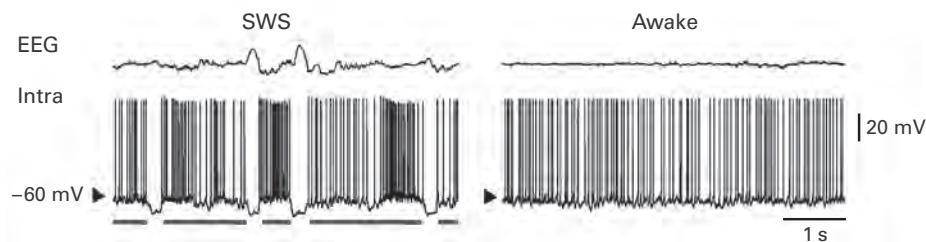
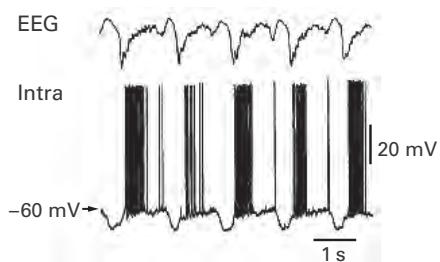
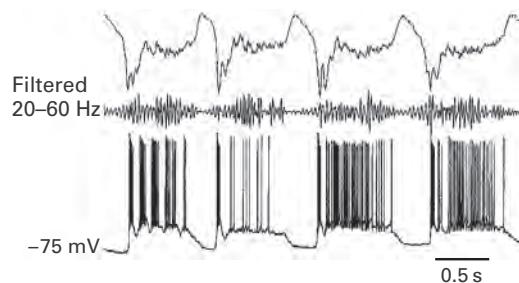
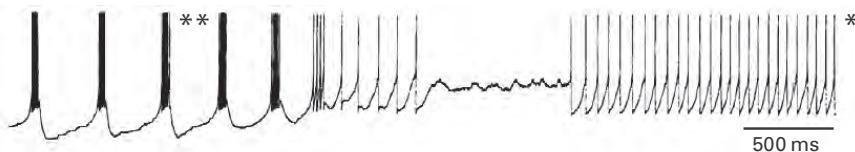
although the early stages of non-REM sleep do not exhibit large slow waves. As non-REM sleep deepens, spindles may appear frequently in the EEG, especially during stage 2 sleep. Spindles are rapidly waxing and waning bursts of EEG activity in the 7–14 Hz range, lasting around 1 second. These oscillatory bursts are nested within slow waves (Destexhe et al., 2007).

In contrast, the scalp EEG of REM sleep resembles wakeful EEG in that traces look irregular or, in other words, nonrhythmic. This resemblance prompted physiologists to label REM sleep alternatively as “paradoxical sleep.” The EEG pattern is also called “desynchronized”¹ to contrast it with the slow-wave patterns of non-REM sleep. However, on closer scrutiny REM-sleep and awake EEG do show bouts of rhythmicity, organized in a much faster range than delta waves (20–80 Hz, peaking in the gamma-band, around 40 Hz; Steriade, McCormick, & Sejnowski, 1993; Steriade, 1996).

Transitions in desynchronized and rhythmic forms of activity are by no means uniform across the whole brain. When we record LFPs inside the hippocampus, for instance in rats, we find that states of attentive wakefulness, orienting, voluntary movement, and REM sleep are marked by a powerful rhythmic activity in the 4–12 Hz range, labeled “theta rhythm” (Vanderwolf, 1969; Buzsaki, 2006). These states are not well-defined and do not provide one-to-one correlates of theta rhythm. Hippocampal signals can be measured using invasive electrodes in animals or using depth electrodes implanted in the brain of epileptic patients undergoing neurosurgery to remove the focus where seizures originate. In contrast to REM sleep, hippocampal LFP signals are irregular during non-REM sleep, except for intermittent high-frequency (140–200 Hz) bursts called “ripples” (see figure 7.3). In area CA1, these ultrafast waxing and waning oscillations correlate to burst firing of cell populations, and to strong synaptic depolarizations of the same cells, reflected in the LFP by

Figure 7.2

Up and Down states in cortical and thalamic neurons. (A) Intracellularly recorded activity of cells in cat cortex during slow-wave sleep (SWS; left) and wakefulness (right). Shown above the intracellular traces are EEG traces recorded with cortical depth electrodes. Slow waves during sleep are paralleled by intracellularly recorded Down (quiet, hyperpolarized) states, whereas the desynchronized EEG in the awake state is paralleled by a continuously depolarized, active state. Dark gray bars in the left-hand panel indicate Up states. (B) Slow oscillations in a thalamocortical cell in cat brain anesthetized with ketamine and xylazine. Downward deflections in EEG oscillations at ~0.75 Hz are paralleled by dense firing (~80 Hz) during the Up state of a neuron in the ventrolateral nucleus of the thalamus. (C) Same as (B), but now an intracellular recording from a thalamic reticular nucleus neuron is shown. EEG has been recorded from motor cortex and is band-pass filtered to highlight activity in the 20–60 Hz range (middle trace). Note how upward EEG deflections are coupled to Down states (hyperpolarized periods). (D) Various neurotransmitters are implicated in regulating the transition from Down to Up states (e.g., noradrenaline, acetylcholine, glutamate, and histamine; Steriade, McCormick, & Sejnowski, 1993). Here, serotonin (5HT) induces a depolarization of thalamic reticular neurons, accompanied by the transition from low-threshold calcium spikes with bursts of action potentials (expanded in D2, ***) to a state of tonic firing of nonbursty spike patterns (D2, *). During the period marked by “–DC” the cell is temporarily hyperpolarized by injection of negative current through the recording pipette, causing a restoration of rhythmic bursting. (A) adapted from Steriade et al. (2001) and Destexhe et al. (2007), with permission from Elsevier. (B, C) adapted from Steriade et al. (1996), with permission from the *Journal of Neuroscience*. (D) adapted from McCormick and Wang (1991), with permission from Wiley.

A Neocortex**B Thalamocortical****C Thalamic reticular****D1 Thalamic reticular****D2**

“sharp waves” and caused by burst discharges of CA3 neurons (O’Keefe & Nadel, 1978; Buzsaki et al., 1992).

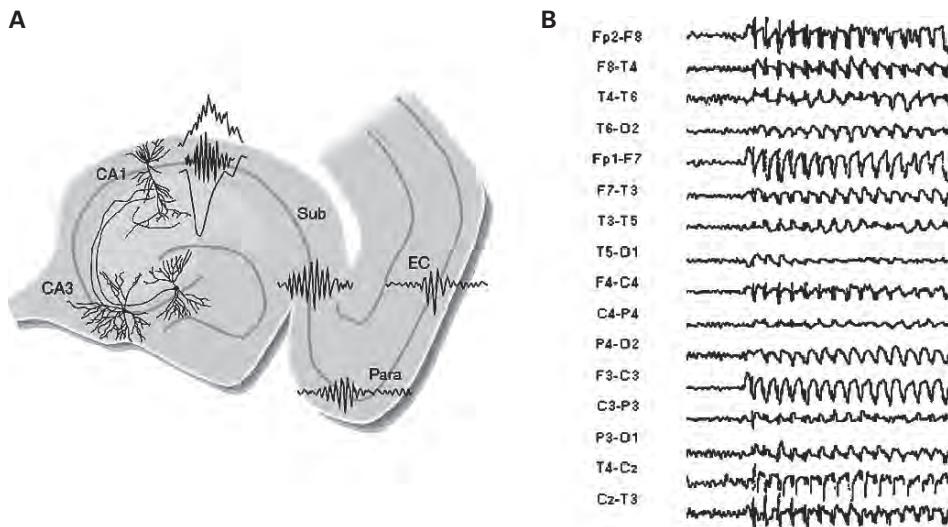
Sleep phases are regulated by a plethora of subcortical areas. Apart from the desynchronized neocortical EEG and hippocampal theta rhythm, REM sleep is characterized by ponto–geniculo–occipital spikes (PGO spikes, also labeled “PGO waves”), which can be recorded from the LGN. These spikes arise because of intense bursts of action potentials fired by cholinergic cells in tiny nuclei in the reticular pons region (Sakai & Jouvet, 1980; Steriade et al., 1990).

The activity of these and other cell groups in the brain stem has been considered an *enabling factor* for consciousness and other aroused states: their activity is probably necessary for bringing a brain system into the appropriate state to sustain conscious experience, yet without contributing *experiential* content (Rees et al., 2002). Cholinergic cells help you to become aware but do not determine what you are aware of. However, acetylcholine release in general not only affects the level of consciousness but also attention, long-term memory, and short-term memory (Broersen et al., 1995; Dalley, Theobald, et al., 2004; Von Linstow Roloff et al., 2007). One of the key actions of acetylcholine in switching the brain from non-REM to REM sleep or wakefulness is to depolarize thalamic reticular nucleus cells as well as relay neurons (see figure 2.5). This prolonged depolarization effectively switches relay neurons to a more permanently active state (Steriade, McCormick, & Sejnowski, 1993; Steriade, 1996; figure 7.2B).

Important as the cholinergic cell groups may be, no single group has been identified that is indispensable for cortical arousal or consciousness (Schiff & Plum, 2000). The cholinergic groups are part of a larger complex of loosely organized nuclei scattered across the medulla, brain stem, basal forebrain, and hypothalamus, collectively forming an ascending arousal system. These nuclei also comprise serotonergic cells in the raphe nuclei, noradrenergic neurons in the locus coeruleus, and histaminergic and hypocretin-releasing cells in the hypothalamus. These neuromodulatory systems show characteristic rising and declining profiles of release across the sleep–wake cycle. Acetylcholine levels are high during waking and REM sleep but low during non-REM sleep. In contrast, serotonin and noradrenalin levels are low in REM sleep but rise during non-REM sleep and waking (Diekelmann & Born, 2010).

7.2 The Brain Stuff That Dreams Are Made Of

Dreams present us with experiences that can be radically different from wakefulness yet belong to the realm of consciousness. When dreaming, we often find ourselves in a quasi-conscious state, in the midst of bizarre scenes and story lines, concocted by our brains for reasons we do not understand well. Dream states are of a hallucinatory nature and have inspired clinical neuroscientists to consider them a “normal delirium” (Hobson & Pace-Schott, 2002). Dreaming resembles a state of self-propelled imagery

**Figure 7.3**

Hippocampal ripple activity and ictal EEG oscillations during absence seizures. (A) Ripples are high-frequency (~150 Hz) oscillations observed in local field potentials recorded from hippocampal regions. Physiological evidence indicates they originate from synchronous bursts of action potentials in area CA3, which propagate to area CA1 via the Schaffer collaterals, from where they are thought to spread across the subiculum (Sub), entorhinal cortex (EC), and parasubiculum (Para). The top and bottom of the picture correspond to the dorsal and ventral side of the hippocampus near its septal pole. Note how ripples recorded from the pyramidal cell layer in CA1 (gray curve) are also manifested in field potentials more dorsally and are paired with “sharp waves” recorded ventrally from the layer. (B) Ictal EEG of a 12-year-old child suffering from absence epilepsy. Originating from the Latin “ictus” (blow), “ictal” refers to the occurrence of an actual seizure. Each row plots EEG traces derived from a pair of electrodes, placed, for example, frontally (F), temporally (T), and so on. The bilateral, near-synchronous onset of spike-and-wave bursts (~3 cycles per second) is clearly visible. The patient was reported to display behavioral automatisms during some of the seizures. (A) from Buzsaki and Chrobak (2005). Reprinted with permission from Macmillan Publishers Ltd. (B) from Yoshinaga et al. (2004), copyright Elsevier.

within wakefulness—except that awake imagery enables us to tell the difference between imagery and external sensory stimulation. Apart from the synthetic nature of dreams, our experience of dreams is distinguished by our inability to direct thoughts logically and to control and plan our in-dream actions voluntarily and by a decreased self-awareness. Marked as dreams often are by delight, joy, or fear, they have an emotional drive that is atypical for normal wakeful experience. Elements from preceding daytime activities may intrude into dreams occurring well into the night, but their content is not greatly influenced by them (Siegel, 2001; Schwartz & Maquet, 2002; Hobson, 2009; but see Wamsley et al., 2010). When subjects in REM sleep are exposed to external stimuli, such as sounds, these are only incidentally incorporated into dreams, as recalled later. Even when objects are visually presented to subjects in REM sleep with their eye lids forced open, these were not reported as having been part of the dream.

Contrary to popular belief, dreaming is not exclusively coupled to REM sleep. Subjects are more likely to recall dreams when they wake up from REM sleep as compared to non-REM sleep, but the incidence of dream recall directly after non-REM sleep ranges between ~10% and 70% (Foulkes, 1962; cf. Oudiette et al., 2012). This recall is not attributable to REM episodes occurring prior to non-REM sleep episodes from which subjects woke up (Suzuki et al., 2004). Non-REM dreams are usually shorter and less bizarre or emotional than reports on REM dreams. The incidence of non-REM dreaming may be higher than subjectively reported because the inability to recall a dream may be due to the absence of dreaming or a disability to store and recall the dream experience. Conversely, the occurrence of REM sleep does not guarantee dreaming. In their original studies, Aserinsky and Kleitman already reported nondreamers showing similar levels of REM sleep as found in dreamers. In Charcot–Wilbrand’s syndrome, focal brain damage causes a loss of dreaming, while REM sleep is preserved (Bischof & Bassetti, 2004).

That both REM and non-REM states can give rise to dreaming is an important lead in identifying “state traits” associated with consciousness. The resemblance between REM sleep and wakeful EEG is clear, but what do REM and non-REM states have in common to explain dreaming? Moreover, why do we not continuously dream during REM or non-REM sleep? Apparently the cortical EEG state is not fully indicative of when, during sleep, we precisely dream and when we do not.

As mentioned, slow-wave activity in the cortical EEG during non-REM sleep is explained by massively synchronized alternations between hyperpolarized (“Down”) and depolarized (“Up”) states of neocortical neurons (see figure 7.2A). The Down state reflects the neuron’s resting level, being devoid of strong excitatory input from neighboring neurons or from the thalamus (Crunelli & Hughes, 2010). The Up state is made up of glutamatergic inputs from many cells, which sum up when arriving nearly synchronously at the postsynaptic dendrites. In this Up state, the membrane potential fluctuates just below or above spiking threshold, so that irregular spike trains are generated. During wakefulness, cortical neurons generate a continuous stream of spikes, whereas slow-wave sleep is marked by Down-state interruptions of otherwise similar spike trains (Destexhe et al., 2007).

Down–Up state alternations in the neocortex are closely coordinated with similar transitions in thalamic relay cells (see figure 7.2A). The relay cells may provide an important excitatory drive to lift cortical cells from a Down to an Up state, although cortical tissue *in vitro* can sustain slow-oscillatory states even in the absence of thalamic input (Le Bon-Jego & Yuste, 2007; Crunelli & Hughes, 2010; Tahvildari et al., 2012). Joining the slow-oscillating orchestra, the neurons of the thalamic reticular nucleus also flip between Down and Up states during non-REM sleep or anesthesia.

The reticular thalamic input to relay cells reinforces rhythmicity in the thalamocortical system, first by inhibiting them but then, as a direct consequence, facilitating a low-threshold Ca^{2+} spike, on top of which Na^+ spikes arise (see figure 7.2D). Nevertheless, this input is unlikely to be the sole pacemaker of slow oscillations. So where *do* the slow

oscillations originate? This may be the wrong question to ask, because rhythmogenesis may be an emergent property of the nonlinear dynamics in distributed systems (Buzsaki, 2006). When we become drowsy and gradually pass into deeper stages of non-REM sleep, the loss of cholinergic tone may first lead to a local and incidental loss of sufficient excitation to maintain a collective and persistent Up state, and next this local breakdown leads to more and more widespread predominance of Down states throughout the system. Conversely, a local coincidence of glutamatergic input and lack of inhibition may boost the system back into an Up state by reverberating and spreading across neighboring thalamocortical microcircuits.

Slow oscillations are not only coordinated within the thalamocortical system, but also more globally, unrolling in step with activity changes in the basal ganglia and hippocampus. By themselves, striatal principal cells will remain stuck in a taciturn Down state unless they receive strong excitatory inputs from neocortical, hippocampal, amygdaloid, or thalamic sources (see chapter 6; Cherubini et al., 1988; Pennartz et al., 1991). Neocortical Down-to-Up state transitions nearly coincide with hippocampal sharp wave–ripple (SWR) activity (Battaglia et al., 2004). In turn, SWR activity is aligned in time with peaks of firing activity in the prime subcortical target of the hippocampus, ventral striatum (Pennartz et al., 2004). It is largely unknown whether phylogenetically old systems such as the brain stem, cerebellum, and hypothalamus play in tune with the forebrain orchestra.

In addition to temporal coordination, slow-wave activity is spatially organized. In neocortical slices maintained in vitro, Sanchez-Vives and McCormick (2000) studied slow-wave activity arising spontaneously in the cortical network. Slow waves traveled from deep to superficial layers, as well as in a horizontal direction at a speed of ~1 cm/s. In EEG recordings, Massimini et al. (2004) also observed a slow oscillation traveling across the human cortex during non-REM sleep, but at a much higher speed (1.2 to 7.0 m/s). The function and mechanism of these frontal-to-caudal waves of activation is unclear but may, speculatively, be related to a spatially ordered reactivation and recall of information stored in neocortical networks, a topic we will be looking at next.

Despite these insights, we do not yet understand why and how non-REM sleep gives rise to dreaming. Reports on subjects woken up during non-REM sleep are rarely specific about the precise EEG state the subjects were in just before they were aroused enough to report dreams. Possibly it was not a “deep sleep” stage from which subjects were woken up to report dreaming; they may have lingered in a more superficial state with irregular EEG activity resembling an awake or hypnagogic state. A second route for further research derives from the shorter duration of non-REM dreams than those associated with REM sleep. During slow-wave sleep, prolonged Up states may give rise to a rapidly self-organizing neural activity corresponding to short dream bouts. In contrast, brief Up states may not allow neural systems to integrate loosely regenerated pieces of information into a coherent representation characteristic of dreaming. Intracellular recordings under anesthesia indicate that Up states may last as long as ~4 seconds (Steriade, Nunez, & Amzica, 1993). This

hypothesis cannot be easily dismissed because dreaming may occur on a different, faster time scale than during wakeful experience. If *all* of our neural activity patterns corresponding to a 1-minute experience were compressed tenfold, we might remain unaware of the difference because our subjective sense of time is “clocked” (or calibrated) on the dynamics of neural activity itself, not on some objective reference outside the brain.

In contrast to non-REM sleep, the relation between REM sleep and dreaming is much more taken for granted. However, if the low-voltage, desynchronized activity of the neocortex and powerful theta rhythm of the hippocampus make REM sleep so similar to wakefulness, why aren’t we simply awake and conscious of our environment in this brain state—despite having our eyes closed? Brain imaging studies have shed some light on this issue. Positron-emission tomography (PET) imaging studies have shown that REM sleep, relative to wakefulness, is marked by an overall decrease in prefrontal metabolic activity (Maquet et al., 1996; Braun et al., 1997). In contrast, the amygdala, parahippocampal cortex, parietal operculum, deep frontal white matter, and anterior cingulate cortex are hyperactive.

The decline in prefrontal activity is reminiscent of the hypoactivity of this area in schizophrenic hallucinations (Andreasen et al., 1997) and may account for the loss of logical, directed thought control and for disinhibition of emotions and sexual arousal during dreaming. Consistent with this, the high metabolic activity in the amygdala and anterior cingulate cortex during REM sleep correlates with the strong emotionality of many dreams, while the parahippocampal cortex may support retrieval of spatial and episodic information and, hence, construction of dream scenes and narrative. High activity of sensory-associative areas such as IT cortex may in part account for imagery and other sensory aspects of dreaming (Braun et al., 1997), consistent with a recent report on consciously exerted auditory and visual imagery (Daselaar, Porat, et al., 2010). Altogether, neuroimaging studies suggest a region-specific pattern of dream activity that is consistent with global scalp-EEG recordings (Hobson, 2009).

What, if anything, can we say about the function of dreaming? Ideas about this are extremely divergent and speculative, ranging from the reemergence of suppressed thoughts and unfulfilled wishes (Freud, 1900), via the internal repetition of instinctive behaviors (Jouvet, 1973), to memory formation and consolidation (Wilson & McNaughton, 1994; Pennartz, Uylings, et al., 2002; Wamsley et al., 2010; Diekelmann & Born, 2010), or to a “protoconsciousness” function serving to develop and maintain waking consciousness (Hobson, 2009). We know that REM sleep contributes to consolidation of memory especially for sensorimotor skills (e.g., Karni & Sagi, 1993), but this does not mean that dreaming per se subserves memory formation. The confabulated nature of dreams, devoid of systematic repetition of preceding daytime experiences, would contradict such a role.

My bet is on the hypothesis that the strong dominance of emotional and memory systems during dreaming, coupled to frontal hypoactivity, drives the brain into an exploration of novel, challenging cognitive-emotional states that enable animals to better cope with

unexpected situations in the future. In case of humans, we may say that dreaming makes us think “out of the box.” The hyperassociative nature of dreaming may enable the brain to explore novel associations between elements that were experienced in isolation from each other in the past, so that novel insights and cognitive strategies emerge (cf. Wagner et al., 2004). Dreams offer us virtual-reality (VR) experiences in which we can freely associate sensations with emotions and spontaneously arising thoughts with unexpected outcomes and links. Normally, a VR exercise would be accompanied by body movement, but fortunately the motor system shuts down during dreaming: it protects us from hurting ourselves and from giving away our shelter position to nocturnal predators.

Experienced contents contrast strongly between dreaming and wakefulness, yet we can say we have conscious experiences in both states and are able to recount them later. The desynchronized nature of the cortical EEG is common to both states, but why would this relate to consciousness? As mentioned, the term “desynchronization” is misleading because this EEG state does permit a waxing and waning occurrence of synchronized firing at a local level and in smaller groups of neurons, and there is evidence that this localized, sparse (rather than massive) synchrony is characteristic of information processing in conscious states (Kuhn, Denk, et al., 2008; Greenberg et al., 2008; Barrett et al., 2012). Leaving the discussion on neural synchrony and binding of information aside for now, local synchrony, particularly in the higher frequency ranges (20–100 Hz), may subserve a gamut of cognitive functions—for instance, perceptual attention, short-term memory, behavioral inhibition, and sensorimotor integration (Murthy & Fetz, 1996a,b; Tallon-Baudry et al., 2001; Van Wingerden et al., 2010; Bosman et al., 2012, 2014).

What the combination of global desynchronization and local, sparse synchrony hints at is that conscious states thrive neither on brain states with *too much* synchrony nor on states having *no synchrony at all*. Optimal for consciousness may be a sparse, intermediate level of synchrony between dispersed groups or assemblies of feature-specific neurons, which flexibly arise and disappear across brief periods of time. Too much synchrony, for instance, occurs during *absence epilepsy* or *petit mal*, marked by a loss of consciousness and of ability to recall afterward what happened during the seizure (see chapter 3). *Absence* is accompanied by massive, highly synchronous thalamocortical rhythmicity, manifested in the cortical EEG as “spike-and-wave” discharges (Meeren et al., 2002; figure 7.3B). In the gamma frequency range, LFP activity is normally associated with perceptual, motor, and cognitive activity, but during epilepsy this 30–100 Hz activity becomes too widespread across thalamocortical systems, too massive, and concomitantly patients become unconscious in this type of seizure (Lopes da Silva, 2004). Why a sufficient level of desynchronization is needed to sustain conscious states is unknown, but a couple of plausible lines can be set out.

First, if synchrony supports cognitive processes and specific area-to-area communications subserving these, it needs to be selectively applied to avoid a situation where all kinds of information demand the same cognitive resources. Excess synchrony would lead to an

abundance of *noise correlations*, meaning that the firing of neurons is correlated by default, due to their preestablished, shared presynaptic inputs, without serving any signaling function in particular (Aertsen et al., 1989; Zohary et al., 1994; Renart et al., 2010; Ecker et al., 2010; Montijn et al., 2014). Highly correlated firing can impair the signaling capacity of populations and hence psychophysical performance (Zohary et al., 1994). Simulations of cortical networks support the plausibility of inhibitory, GABAergic mechanisms that “track” the excitatory inputs to cortical neurons closely in time and function to *decorrelate* the activity of their members (Renart et al., 2010). Second, the firing rate of a neuron may not be the only way by which it codes information. The *phase* at which a cell fires relative to another member (or to an LFP oscillation) provides an alternative means (O’Keefe & Recce, 1993; Pennartz, 2009; Nadasdy, 2010), and this option for *phase coding* will be impaired when neural synchrony becomes excessive.

As a final caveat about REM sleep, let me emphasize that it can occur without accompanying dreams. Apparently, a desynchronized cortical EEG state is *insufficient* to characterize consciousness. Conscious experiences do not come out of the blue while the cortex is lingering in this state. Further principles for organizing and structuring them are necessary, and these will be pursued below.

7.3 Replay: The Return of Experiential Information during Sleep

One of the most perplexing discoveries in the field of sleep physiology over the past 2 decades has been the phenomenon of reactivation or replay. Following initial observations by Pavlides and Winson (1989), Wilson and McNaughton (1994) recorded firing activity from groups of neurons (“ensembles”)² in the hippocampus of rats that subsequently engaged in a first episode of sleeping, running a maze, and a second sleep episode. While the rats ran the maze repetitively, hippocampal place cells were activated in a stereotypic sequence, marked by specific temporal correlations between their firing patterns. This groupwise pattern of correlations specifically returned during non-REM sleep, whereas it was not seen during the prerule sleep session. The term “replay” is justified because the place-cell sequences during postrule sleep preserved the temporal order of events as they had occurred during maze running (Skaggs & McNaughton, 1996; Lee & Wilson, 2002).

Because of the vital role of the hippocampus in episodic memory formation (Scoville & Milner, 1957; Eichenbaum et al., 2011), replay of place-cell information patterns was proposed to support memory consolidation, that is, the gradual strengthening of memories that occurs once the real-life experience is over (Marr, 1971; Wilson & McNaughton, 1994; McClelland et al., 1995; McGaugh, 2000). However, replay is not limited to the hippocampal memory system, as other brain structures such as the ventral striatum, prefrontal cortex, and other neocortical areas exhibit the phenomenon as well (Hoffman & McNaughton, 2002; Pennartz et al., 2004; Euston et al., 2007; Ji & Wilson, 2007). For instance, ventral striatal neurons change firing rate when an animal expects reward and performs behavior

toward reward sites—and it is this type of information that is reactivated during non-REM sleep (Lansink et al., 2008). In a similar vein, the medial prefrontal cortex replays rule-related activity during sleep (Peyrache et al., 2009). Thus, brain areas mostly reactivate the kind of information they were already dedicated to during awake experience. Despite their low resolution, neuroimaging (functional magnetic resonance imaging; fMRI) and EEG data in humans are consistent with reactivation-like phenomena in the awake-resting or sleeping brain (Peigneux et al., 2006; Rasch et al., 2007; Daselaar, Huijbers, et al., 2010; Tamminen et al., 2010).

When does replay precisely occur during sleep, and is it associated with any brain state in particular? Accumulated results have now firmly demonstrated reactivation during non-REM sleep and especially during hippocampal SWR activity (Kudrimoti et al., 1999; Pennartz et al., 2004; Lansink et al., 2008; Peyrache et al., 2009; Davidson et al., 2009; Jadhav et al., 2012; figure 7.4, plate 4). Reactivation during REM sleep has been reported (Louie & Wilson, 2001) but was not found prominently in several other studies (Kudrimoti et al., 1999; Lansink et al., 2008, 2009; Tatsuno et al., 2006). Neural ensembles do display distinct activity patterns during REM sleep, but it is unknown whether these patterns reflect spontaneously generated concoctions or contain elements from preceding experiences in less conventional ways than studied so far (cf. Poe et al., 2000; Grosmark et al., 2012).

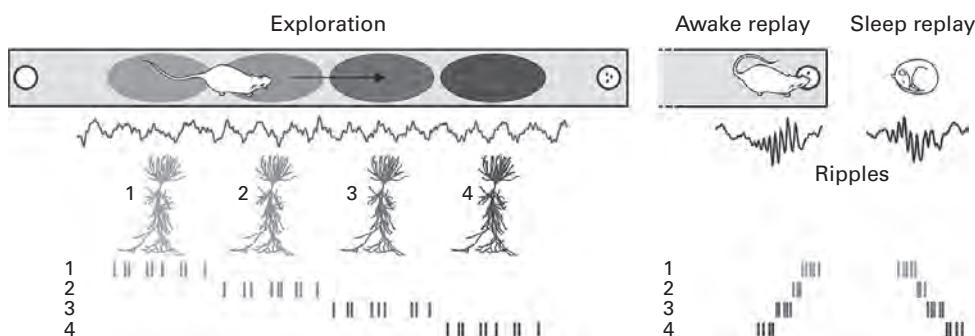


Figure 7.4 (plate 4)

Schematic of neuronal replay associated with sharp-wave ripple activity in the hippocampus. The top left panel shows a rat running along a linear track toward a reward site at the end of the track. The trace below the track represents theta rhythm in the local field potential recorded from hippocampus during running. Below this trace, four hippocampal pyramidal neurons (numbered 1–4) are shown, each firing spikes (color coded, vertical ticks) at subsequent places on the linear track. Thus, running is associated with a sequence of place-cell firing patterns in hippocampus (lower panel). Right-hand panels: When the rat arrives at the reward site and consumes food, ripples are generated in the hippocampal EEG, during which the place-cell sequence reactivates in a temporal order opposite to that of the running sequence (reverse replay). During subsequent slow-wave sleep, place cells reactivate in the same order as during track running (forward replay). From Girardeau and Zugaro (2011), with permission from Elsevier.

Because ripples are “flashes” made up of burst-like, rhythmic population activity, lasting roughly 50–100 ms, one may wonder how much information could be packaged in such brief epochs. At least in three different brain structures—hippocampus, prefrontal cortex, and striatum—spike sequences appear to be 7–10 times temporally compressed during replay, compared to the pace at which firing-rate changes unfold during the original experience (Nadasdy et al., 1999; Euston et al., 2007; Lansink et al., 2009). A single ripple of ~100 ms can thus harbor a spike sequence capturing roughly 1 second of behavioral experience, which is not that much if the animal needs to memorize a longer chain of events. However, Davidson et al. (2009) described how replay of a more prolonged experience can occur across a series of ripple events.

Does replay exclusively occur during sleep, or is it also found in the awake state? Examining rats that ran back and forth across a linear track with reward sites at either end, replay was found to occur during wakefulness, but comes, surprisingly, in two varieties (Foster & Wilson, 2006; Diba & Buzsaki, 2007). Before the rat departed from a reward site to undertake the next run, *forward replay* took place: hippocampal cells fired in the same temporal order as during the upcoming run toward the other end. However, when rats completed their run and paused—for instance, to consume a reward—*reverse replay* was seen (see figure 7.4, plate 4). Now, the temporal order was opposite to the sequence the place cells were going through when the rat was on its way toward that site. During reverse replay, cells that were most recently activated during the approach to the site fire first, followed by cells that were active at progressively earlier moments during the run. Metaphorically, the “movie” would be played back in reverse mode, with actors appearing to walk backward. As for forward replay during sleep, reverse replay likely reflects more than a simple, straightforward reproduction of experienced activity patterns. For instance, a rat’s hippocampus can construct sequential neural patterns corresponding to routes the animal never experienced before (Gupta et al., 2010). Such “creative” concatenations could not be explained from random combinations of previous track experiences. These and related findings indicate that replay may not be an exact reproduction of previous experience, but rather a constructive process in which experiential information may be used to complete a representation of possible paths through an environment. As such, replay at choice points and prior to locomotion to goal sites may reflect planning and internal simulation (Pfeiffer & Foster, 2013; Pezzulo et al., 2014).

Should we think of ripples as reflecting isolated replay events, disconnected from other memory operations going on elsewhere in the brain? There is strong evidence this is not the case. Brain structures show a temporal coordination of replay events, suggesting that the various modalities or “pieces of information” belonging to an experience are jointly reprocessed, close together in time. Experiential “signatures,” specified by correlations in multineuron firing patterns, have been demonstrated not only within single structures such as hippocampus but also *between* structures—for instance, between hippocampus and neocortex (Ji & Wilson, 2007; cf. Hoffman & McNaughton, 2002).

Theories about memory formation posit that it should be the hippocampus that initiates the replay during sleep, and that its recurring activity patterns are propagated via output pathways to target areas, where subsequent replay will be orchestrated (Marr, 1971; Buzsaki, 1989; McClelland et al., 1995; Battaglia et al., 2011). A suitable pathway to test this prediction is the hippocampal projection to the ventral striatum, especially because it is unidirectional. In line with the prediction, Lansink et al. (2009) found that these two structures reactivate together, with the hippocampus kicking off replay events in most instances, followed by ventral striatum. As further evidence for temporal coordination of replay events across multiple brain structures, Siapas and Wilson (1998) showed that hippocampal ripples correlate in time with spindle activity recorded in the medial prefrontal cortex (see also Wierzynski et al., 2009).

7.4 Replay as a Case of Nonconscious Representation

Replay currently figures as a model for memory consolidation but is no less interesting from the viewpoint of consciousness. Replay is marked by highly structured activity of selective neural groups, and this property has also been proposed as a hallmark of conscious experience (Tononi & Edelman, 1998; Crick & Koch, 2003; Edelman et al., 2011). Could we be aware of events coded by replayed neural activity, or is this activity too dissimilar from patterns typically associated with consciousness?

At first glance, replay does not seem to have the right properties to lend itself to conscious processing. For example, replay events are brief and appear as “snapshots” rather than a continuous stream of consciousness. However, this argument can be countered by the observation that longer stretches of experience can be captured by consecutive ripple-replay events segregated by silent intervals (Davidson et al., 2009). Silent intervals do not necessarily pose a problem because a break, or temporal discontinuity, would only become part of our experience when there are cells that code the discontinuity—but also these cells would go silent, if present.

A stronger case against replay as a conscious process is presented by backward replay. Suppose we neglect the temporal compression characteristic of replay and consider what it would be like to experience a hippocampal replay event as a sequence of spatial positions we occupy. In backward replay, we would see ourselves “walking backward” while sipping from a rewarding drink. One might object that a full backward replay in the brain should also set the brain’s time-tracking mechanism in reverse mode, so that we would not notice the difference with forward replay. However, the argument leans on our common experience of being at a single place at a given experiential moment, and of walking as a progressive change in spatial location from *A* to *B* to *C*. Backward replay should then result in an experienced order of *C* to *B* to *A*, which is an experience alien to us unless we put visual imagery at work.

A further argument against replay as a consciously experienced event holds that the main sleep phase exhibiting replay is non-REM sleep, whereas quasi-conscious states (i.e., dreams) are more frequently reported when people wake up from REM sleep. Although replay events may sometimes share a synthetic, constructive property (Gupta et al., 2010) with dreaming, we identify and quantify replay by its strong experiential, repetitive component—unlike dreams which are manifest as confabulatory constructs, occasionally flavored by an intrusive element from previous experience. These arguments by no means close the discussion, but the current evidence argues in favor of replay as a phenomenon during which experiential information is represented in a structured but nonconscious way.

This has several implications for theories of consciousness. In their theory of visual awareness, for instance, Crick and Koch (2003) argue that conscious experiences correlate with selective activity patterns in brain areas engaged in representing features of visual scenes, for example, movement represented by neuronal subsets in cortical area MT and facial shape by IT cortex neurons. Neurons in such areas would forge a “coalition” with each other, meaning that the cortex harbors cell groups that may form cooperating assemblies that rival other groups. This competitive element embodies selective attention and dominance in conscious perception, as illustrated by binocular rivalry (Crick & Koch, 2003):³

The various neurons in a coalition in some sense support one another, either directly or indirectly, by increasing the activity of their fellow members. The dynamics of coalitions are not simple. In general, at any moment the winning coalition is somewhat sustained, and embodies what we are conscious of. (p. 121)

Attractive as this element of their framework may be, the features of replay force us to rethink whether coalitions might be sufficient to explain consciousness. Cross-structural replay shows that each brain area contains cell groups that are selectively activated and also form a “coalition” in the sense that the replay pattern in area X specifically arises in conjunction with replaying cells in area Y, and apparently not with less favored assemblies. The brain has globally operating mechanisms ensuring the temporal alignment of structured activity, as evidenced by the correlations between SWRs and replay in hippocampus, spindles and Up-state transitions in neocortex, and surges of replay in striatum. Altogether this appears to make a pretty strong coalition! As most replay events are unlikely to surface in conscious experience either in humans (as ripple prevalence in human medial temporal lobe is high during non-REM sleep; Axmacher et al., 2008) or other animals, the concept of “coalitions” is not necessarily invalid but seems insufficient to account for the conscious–nonconscious dichotomy. A more refined description of assembly characteristics during conscious versus nonconscious states appears needed. The replay argument applies equally to other frameworks for consciousness, such as global workspace theory. Long-distance broadcasting of sensory information and exchange between higher associational cortical areas is a key hallmark of this theory, but these elements may well be shared by globally coordinated replay (cf. Hoffman & McNaughton, 2002).

If we now return to the finding that non-REM sleep does not accommodate much dreaming activity, we are yet facing a different question: why would conscious representation fall apart during the bulk of non-REM sleep episodes, including widely occurring replay events? As yet, there is no easy solution to this question, but the short-lived nature of replay events may provide a hint. Replay probably involves a time- and space-restricted recruitment of cells across various brain structures. This restricted nature allows cell-to-cell specific correlations to return during sleep, which supports memory consolidation but may be too brief to let multi-area assemblies self-organize into a state in which many sensorial elements can be integrated into a full-blown, multimodal representation.

7.5 Pharmacologically Induced Loss of Consciousness—Anesthesia

Anesthesia is unique in that it is a purposefully induced nonconscious state, aiming to avoid acute pain and discomfort but leaving brain function otherwise intact as much as possible. As in deep sleep, sensory-evoked EEG responses of the primary sensory cortices remain essentially present during general anesthesia although they may be modified depending on the agent used (Sloan, 1998; Supp et al., 2011), suggesting that the relaying of sensory information from periphery to neocortex still works. So what aspect of systems functioning explains that consciousness is lost? Here we will contemplate how brain states under anesthesia differ from wakeful conditions and consider how this compares with insights from sleep physiology and neurology.

There is a great variety of anesthetics, and it is remarkable that they also work by a variety of neural mechanisms to induce the same global effect, the temporary abolition of consciousness. Anesthetics are commonly subdivided according to their gaseous or liquid forms. Volatile compounds include, for instance, nitrous oxide (“laughing gas”) and isoflurane, whereas compounds such as propofol, ketamine, and barbiturates are intravenously applied. Not surprisingly, the general upshot of anesthetics is to shift the overall balance of excitation and inhibition in the central nervous system toward less excitation and more inhibition (but see Haider et al., 2013). A common denominator of both intravenous and inhalational compounds is that they augment GABA_A receptor function, which is mediated by the opening of chloride-conducting ion channels (see chapter 2; Alkire et al., 2008; Franks, 2008). Some anesthetics, including ketamine and nitrous oxide, reinforce GABA_A receptor action mildly, but if this is the case they have additional effects such as antagonism of glutamate receptors (NMDA, AMPA, and/or kainate receptors; all excitatory). A powerful combination found in other agents is that they strengthen the action of both GABA_A receptors and potassium channels, many of which keep the neuron quiet and hyperpolarized when active. A host of effects on other transmitter systems and ion channels has been reported. Notable is that many anesthetics inhibit acetylcholine receptor function (Franks, 2008; Alkire et al., 2008; Forman & Miller, 2011; both nicotinic and muscarinic receptor antagonism has been reported although not necessarily both by the same agent).

This inhibiting effect fits the role of cholinergic and other modulatory cell groups in the brain stem and basal forebrain as important enabling factors for wakefulness and attention.

At cellular and mesoscopic levels, the physiological effects of anesthetics are often similar to the progressive changes seen during non-REM sleep. Starting in the awake state, anesthetics cause a gradual emergence of slow oscillations, made up of alternations between the Up and Down state, with a periodicity of about 1 second. Deeper anesthesia prolongs the Down state and compresses the active state to bursts of cortical activity, expressed in the EEG as a “burst-suppression” pattern (Sloan, 1998; Ching et al., 2012). In the context of non-REM sleep, I argued that brief surges of Up-state activity can harbor replay events but are unlikely to coincide with dreaming or other conscious activity. Here again, the presence of activity bouts under anesthesia confirms the conclusion drawn earlier: specific neural activity per se is not sufficient for consciousness, and consciousness is likely dependent on the progressive development of self-organizing activity across time and a particular way of *structuring* information flow.

Recording ensembles from rat auditory and somatosensory cortex, Luczak et al. (2009) found specific sequences of multineuron firing in response to tones of variable frequency. These sequences were fairly stereotyped and unfolded across ~100 ms. Because stereotyped tone-specific sequences occurred during both wakefulness and anesthesia, they are not a hallmark of consciousness per se: temporally structured activity does not pull off the trick. Overall, experimental results make the question of what distinguishes wakeful versus nonconscious sensory processing even more pressing than it already was.

In vivo two-photon microscopy studies are shedding light on the matter. Using Ca^{2+} sensitive dyes that make cortical neurons fluorescent, spatiotemporal patterns of cortical activity can be studied in large groups of identifiable, single neurons. Greenberg et al. (2008) and Goltstein et al. (2011; see figure 7.5, plate 5) found that the anesthetized state was marked by higher and less variable cell-to-cell correlations in the spontaneous activity of neocortical cells than awake-state correlations. As noted for slow-wave sleep, synchrony or high correlation per se is not conducive to consciousness. On the contrary, the collective results suggest that multineuron activity should be sufficiently *decorrelated*, probably to allow a situation in which only selected cellular subsets can engage in experience-specific correlational patterns. The correlational structure of cortical activity under spontaneous or poststimulus conditions offers a good starting point to investigate more deeply how spatio-temporal activity patterns are constrained to permit conscious processing.

Which sites in the brain are most critical for loss of consciousness induced by anesthesia? In addition to the neocortex, the thalamus has often been designated as an accomplice, because its metabolic demands and blood flow go down as people pass the threshold of unconsciousness (Alkire et al., 2007, 2008). Logical as this thought may be, it is not fully established that the thalamus would always need to be intact to maintain wakefulness. Even after large thalamic lesions, cortical tissue can sustain an activated EEG state. When, for instance, the ventroposterior thalamus is lesioned and somatosensory information is not

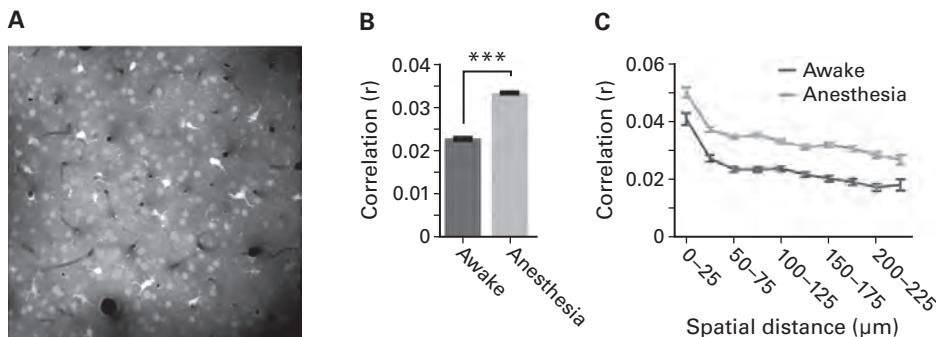


Figure 7.5 (plate 5)

Two-photon imaging of neuronal ensemble activity in the awake and anesthetized visual cortex of the mouse. (A) Two-photon microscopy permits researchers to obtain images with sharply delineated neuronal somata (green objects) and glial cells (specifically, astrocytes; yellow objects) situated in the superficial layers of mouse visual cortex. Green fluorescent staining is achieved by loading the tissue with a Ca^{2+} -sensitive dye, Calcium Oregon Green BAPTA-AM. Spiking activity of individual neurons leads to a rise in intracellular calcium concentration, which is coupled to a transient enhancement of emitted green light. (B) When the fluorescence signals of individual cells are recorded over time, spontaneous fluctuations in electrical activity of the cells result in a pattern marked by transient peaks (firing activity) alternating with quiet periods. When cell-to-cell correlations are calculated across all pairs, studied both during wakefulness and isoflurane anesthesia, a higher correlation (shown as mean \pm s.e.m.) is obtained for the anesthetized condition (***, $P < 10^{-10}$, Wilcoxon's matched pairs signed rank test). (C) This correlation tapers off as the cortical distance between the members of a pair increases, and the difference between awake and anesthetized conditions is relatively constant across distance. From Goltstein et al. (2011; cf. Goltstein et al., 2013).

relayed to the neocortex—resulting in hypesthesia of the hand—the sensorimotor cortex is metabolically activated as in control subjects upon hand vibration (Remy et al., 1999). Even after massive, bilateral thalamic lesions, the cortex may sustain an activated state (Villablanca & Salinas-Zeballos, 1972; Vanderwolf & Stewart, 1988). Devoid of thalamic input, cortical cell groups maintained *in vitro* are able to generate spontaneous patterns under pharmacologically controlled conditions (Cossart et al., 2003; Crunelli & Hughes, 2010). Probably the neocortex contains its own circuit-level mechanisms for sustaining spontaneous activity in the presence of enabling factors such as acetylcholine, but these mechanisms are strongly facilitated by the oscillatory engines of the thalamus.

That thalamic metabolism is reduced upon loss of consciousness may be explained from its level of excitation strongly depending on output from cortical L5 and L6 neurons (Vahle-Hinz et al., 2007; Blethyn et al., 2006; figure 6.2). When Parkinson patients undergo a surgical implantation of deep-brain stimulation electrodes inside or close to the subthalamic nucleus, the effects of anesthesia on both their cortical and subcortical EEG can be studied. Once they lose behavioral signs of consciousness, the cortical EEG changes simultaneously. In contrast, the “wakeful” pattern of subcortical EEG signals—which probably reflect thalamic activity to a great extent—does not change until minutes later into the period of anesthesia (see figure 7.6; Velly et al., 2007). Thus, whereas the neocor-

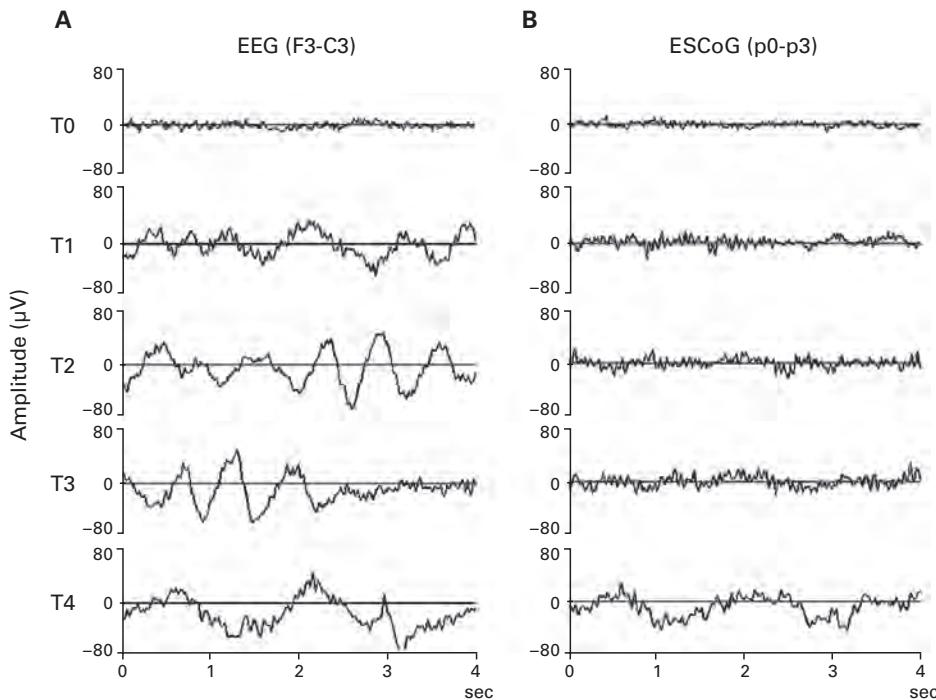


Figure 7.6

Cortical and subcortical field potential recordings during induction of anesthesia in patients with Parkinson's disease. Traces in left-hand panel ("EEG") are representative examples during wakefulness (T0), shortly after loss of consciousness induced by propofol (T1), before laryngoscopy (T2), following intubation (T3), and 5 minutes later (T4). Consciousness was assessed at 20-second intervals by the patient's response to a loud voice command ("Open your eyes") with mild tapping on shoulder. F3–C3 refer to EEG recording positions largely over the frontal cortex. ESCoG in the right-hand panel refers to subcortical field recordings from the subthalamic nucleus region. Leads p0–p3 are ordered dorsoventrally on quadripolar deep-brain stimulation electrodes. Whereas the appearance of cortical delta waves correlates tightly with loss of consciousness (T1), subcortical delta waves only emerge after intubation (T4). Adapted from Velly et al. (2007).

tex remains the prime suspect in explaining anesthesia-induced loss of consciousness, the exact role of the thalamus is less certain.

Some thalamic nuclei appear more important than others. Lesions of relay nuclei will prevent conscious sensations of inputs in the corresponding modalities (cf. Schmid et al., 2010) but do not abolish cortical arousal altogether or experiences in other modalities. In contrast, the intralaminar thalamic nuclei are thought to exert a more generalized influence on arousal and consciousness (Llinás & Ribary, 2001; Schiff et al., 2002; Vogt & Laureys, 2005). As these cell groups project to superficial cortical layers, they may modulate integrative, corticocortical communication via this route (see figures 6.2 and 6.11) but in addition subserve other functions, such as modulation of striatal excitability (Groenewegen & Berendse, 1994; Jones, 2002). Thirdly, the higher-order relay nuclei of the thalamus, such

as the pulvinar and mediodorsal nucleus, may well be needed to pace or drive communication between primary and higher cortical areas (Guillery & Sherman, 2002; Sherman, 2006).

Given the complex involvement of different thalamic nuclei, is it possible that some parts of the neocortex remain “awake” under general anesthesia, whereas others fall silent? In rare instances (0.1–0.2% of the population), patients have been inferred to reach some level of consciousness during surgeries where general anesthesia was assumed, as they were unable to respond to verbal commands or other external stimuli. After recovery, they are able to verbally report on their experiences (Myles et al., 2004; cf. Noreika et al., 2011). Ketamine can incapacitate a patient’s decision-making abilities and motivation to respond to questions (Alkire & Miller, 2005). The patient may have conscious feelings, but is unable to communicate them. Conversely, when using some anesthetics, the frontal cortical areas do not become inactive, while caudal brain areas associated with sensory processing are deactivated, which is in line with the point made earlier (see chapter 6) that frontal cortical activity does not seem essential for consciousness in general. Thiopental, for instance, deactivates posterior cortical areas but affects activity in frontal cortex much less, even at doses inducing hypnotic effects (Veselis et al., 2004; an exception being the middle frontal gyrus; see figure 7.7, plate 6). Nevertheless, one must be careful here because an area’s activity is usually defined by global metabolism and blood flow, whereas also its mode of activity can change (cf. Supp et al., 2011).

These posterior areas form a group that appears more generally deactivated by anesthesia, as assessed by metabolic measures, and may constitute core components needed for consciousness (see figure 7.7, plate 6; Kaisti et al., 2003; Veselis et al., 2004; Vogt & Laureys, 2005; Franks, 2008; Alkire et al., 2008). They comprise the cuneus and precuneus, various temporal-lobe areas, posterior cingulate cortex, inferior parietal cortex, and temporo-parietal junction area. Some of these areas are members of the “default mode network,” a group of brain regions especially active at wakeful rest and implicated in cognitive operations such as memory recall, self-monitoring, self-awareness, and imagery (Raichle & Snyder, 2007; Buckner et al., 2008; Huijbers et al., 2012). The temporo-parieto-occipital junction, lying at the tripoint where the main posterior lobes meet, is a higher cortical area integrating inputs from memory and the visual, haptic, and vestibular senses (Amedi et al., 2001). Altogether, these posterior areas appear to be involved in integrating information from different sensory modalities and may integrate information recalled from memory with perceptual input, making them suitable candidates for contributing to multimodal representations as a core feature of consciousness.

Loss of consciousness may not only, and not simply, involve the deactivation of single cortical areas. Ketamine, for instance, increases cerebral blood flow in many regions, and this is generally associated with elevated neural activity (Langsjo et al., 2005). A study by Ferrarelli et al. (2010) suggests that the functional connectivity⁴ of cortical areas may also be an important parameter. Comparing wakefulness and midazolam-induced

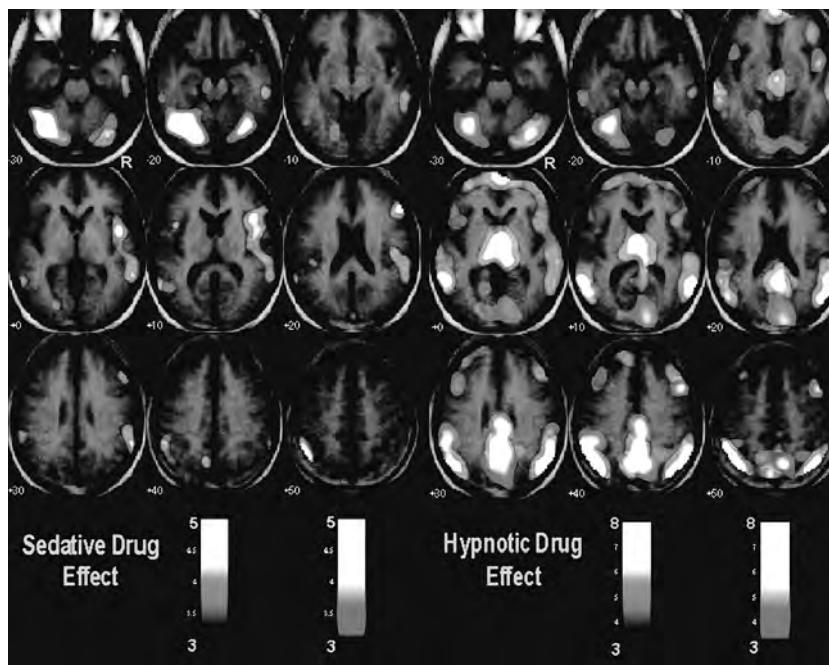


Figure 7.7 (plate 6)

The anesthetics propofol and thiopental affect different brain regions as assessed by induced decrements in regional cerebral blood flow (rCBF). Drug effects at sedative doses (with subjects being responsive to voices or light touch) are visualized in the left-hand subset of nine brain images; drug effects at hypnotic doses (with subjects unresponsive) are on the right-hand side. Propofol and thiopental effects are rendered by a hotness (white–red) and coolness (white–blue) scale, respectively. Horizontal brain sections are shown, going from ventral (top left) toward dorsal slices (bottom right). Focusing on hypnotic effects, thiopental generally induces stronger decreases in posterior cortical areas, whereas propofol decreases activity in both posterior and more frontal regions. The two drugs also share a subset of commonly affected areas (including precuneus, cingulate cortex, inferior parietal lobule, middle frontal gyrus, middle temporal gyrus, supramarginal gyrus, and superior temporal gyrus). Some of the areas affected by hypnosis also show a decreased rCBF under sedation, when subjects are conscious but additional side effects have been documented (e.g., amnesia). From Veselis et al. (2004), with permission from Lippincott Williams Wilkins–Kluwer.

anesthesia, Ferrarelli and colleagues recorded cortical EEG responses to local stimuli applied to the human cortex by transcranial magnetic stimulation (TMS; see figure 7.8, plate 7). In the conscious condition, TMS pulses not only activated the cortical spot that was directly affected (premotor cortex in this case) but also triggered a pattern of EEG waves across other, sometimes distant cortical areas, often lasting ~300 ms after the TMS pulse. In contrast, under anesthesia TMS evoked only short-lasting EEG responses in the cortical neighborhood of the stimulated area, fading out within 150 ms. This difference resembles the result from a similar TMS–EEG experiment carried out during wakefulness versus non-REM sleep (Massimini et al., 2005). In agreement with Luczak et al. (2009), these results prompt us to study the conscious–nonconscious dichotomy beyond a local

level of cortical processing—to scrutinize the patterns of interaction between cortical areas and their ensembles across short and long distances.

7.6 Conscious Processing on Fine Time Scales

So far, we have been considering brain states lasting from several minutes to hours. However, the contents of our consciousness can shift more rapidly. We can quickly switch between a state of daydreaming to alertness when the phone starts ringing. We may stare at a patch of woodland soil covered with leaves and not see anything in particular, until—suddenly—a spider stands out. What happens in the brain between the time a stimulus is presented and the moment we see it?

A visual stimulus triggers a wave of activity traveling from the retina to primary visual cortex in ~35 ms and hence to higher visual areas (cf. figure 6.3, plate 3). When subjects are shown two different types of stimulus and are asked to decide whether an image contains an animal, scalp EEG recordings of event-related potentials (ERPs) show the same cortical stimulus-evoked responses within 150 ms after stimulus onset. However, these responses develop differently for decisions on animals and nonanimals after this time point (Thorpe et al., 1996). At this time, the input has activated shape-sensitive regions, such as IT, and additional regions involved in interpreting the input statistics and responding to it, such as frontal cortex but also lateral occipital cortex (Scholte et al., 2009). When animal or non-animal pictures are only briefly shown and then replaced right away by another image—a so-called “backward mask,” referring to the temporal order of presenting the mask after, and not before, the target stimulus—subjects are still successful in identifying the nature of the stimulus but do not become conscious of having seen an animal, or anything else for that matter. The mask does not prevent the “forward sweep” from activating higher brain systems to discriminate and respond to the stimulus, but it does prevent awareness (Lamme & Roelfsema, 2000). This effect even persists when the interval between stimulus and mask rises up to 200 ms. Apparently, a percept takes time to be constructed and remains susceptible to disruption within this time span.

This result is in agreement with the concept of “microgenesis.” This holds that the neural processes need a slice of time to unfold or self-organize in such a way that percepts can be established (Calis et al., 1984; Calis, 1984; Bachmann, 2006). Backward masking can be thought of as triggering a “race” within the brain, with the visual information of the target stimulus (e.g., an animal picture) in a lead position and activating temporal, parietal, and frontal cortex before the wave carrying information of the masking stimulus arrives. But which pattern of activity in this system determines whether we will become aware of the target stimulus?

This question was pursued by DeHaene et al. (2001), working with subjects viewing streams of images inside an MRI scanner. They studied differences in blood oxygenation level-dependent (BOLD) signals⁵ as well as ERPs in scalp EEG between

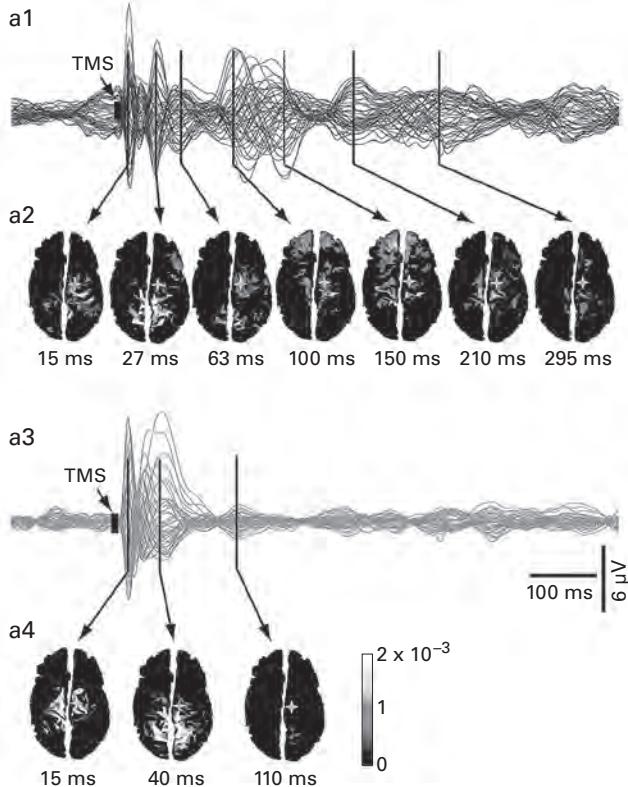
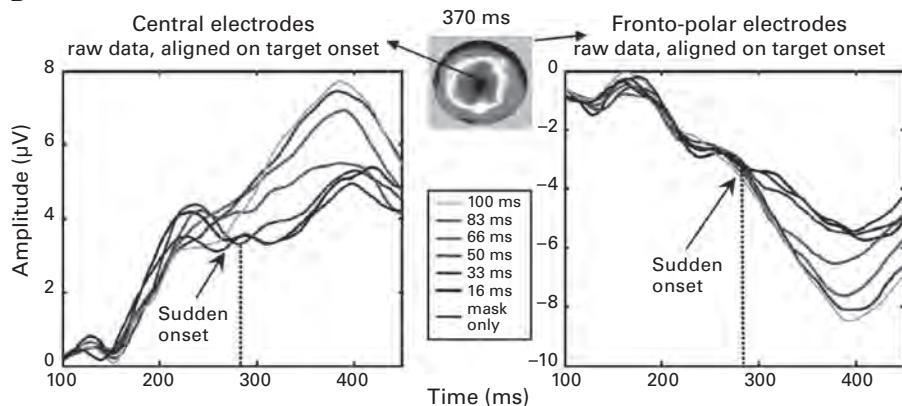
A**B**

Figure 7.8 (plate 7)

Neurophysiological studies on consciousness using anesthesia and stimulus masking paradigms. (A) Anesthesia induced by midazolam shows how scalp-EEG responses evoked by transcranial magnetic stimulation (TMS) of the premotor cortex are altered upon loss of consciousness; a1 and a3 show superimposed, averaged TMS-evoked EEG patterns for wakefulness and anesthesia, respectively. Using high-density EEG mapping, local cortical currents can be estimated (a2: wakefulness; a4: anesthesia). Wakefulness and anesthesia differ in that wakeful EEG responses are more long lasting (> 300 ms poststimulus) and occur over more distributed areas, whereas anesthesia is marked by high-amplitude, short (< 150 ms) responses that remain confined in cortical space. (B) When awake, human subjects are briefly exposed to a stimulus that is followed by a backward mask, their perception of the stimulus depends on stimulus onset asynchrony (SOA; the time interval between target stimulus and mask). In this experimental setup the threshold for reported percepts is about 50 ms. Event-related potentials (ERPs) are recorded at high density from the scalp. Left and right graphs show ERPs from central electrodes and frontopolar electrodes, respectively. Around 370 ms after stimulus onset, the central recording sites show a large peak in amplitude (left), the so-called P3 waveform. The voltage map in the middle illustrates how this waveform is spatially distributed across the scalp, using a SOA of 100 ms. When the SOA is varied from 16 to 100 ms, the central electrodes display a nonlinear increase in EEG amplitude at a latency of about 270–300 ms. The onset of this increment is more sharply aligned in time on the frontopolar electrodes. This nonlinear event can be correlated to a similar increase in subjectively reported perceptibility and has been interpreted as indicating a late “ignition” of distributed activity in fronto-parietal-temporal networks as a correlate of conscious reportability. (A) from Ferrarelli et al. (2010), with permission from the National Academy of Sciences (U.S.A.). (B) from Del Cul et al. (2007).

visual presentations in which words were visible and ones in which they were preceded and followed by masking stimuli, precluding conscious vision. Word presentations lasted for only 29 ms. Both seen and unseen stimuli reliably activated the areas in the left hemisphere, including the fusiform gyrus, extrastriate visual cortex, and frontal areas, which are known to take part in shape perception and word reading, but only the “seen” condition was accompanied by a widespread activation of prefrontal and parietal areas (DeHaene et al., 2001). DeHaene and colleagues concluded that a frontoparietal network is specifically engaged in the ability to report consciously seen words.

Still, the riddle remains how a backward mask could prevent a reported, conscious percept, because the higher parietal-frontal areas will be activated before the mask information arrives there. One solution to the problem holds that, at some level of the visual hierarchy, recurrent activity via corticocortical or corticothalamic projections must occur for a conscious percept to arise (Churchland, 1995; Allman et al., 1985; Lamme et al., 1998). According to DeHaene et al. (2006), signals will be “broadcast” across the parietal-frontal network and connected areas forming part of the hypothesized global workspace. In both scenarios, a forward sweep conveying mask information will interfere with the processing needed to achieve perceptual genesis. When liberally applied, the broadcasting concept of DeHaene and colleagues is compatible with perceptual theories advocating reciprocal signaling between areas, either low or high in a hierarchy.

Del Cul et al. (2007) followed up these masking experiments with high-density scalp recordings of ERPs (see figure 7.8B, plate 7). They varied the delay between target stimulus and backward mask in small time steps. Would it be possible to show a temporally discrete moment, just after stimulus presentation, where an electrophysiological correlate of the hypothesized broadcasting process would become manifest? If so, this would

correspond to subjects' reporting a rather sharp jump from "unseen" to "seen" responses to stimuli as the delay is prolonged. Recordings showed an early stage of response activity in occipitotemporal networks occurring within 250 ms of stimulus onset, consistent with processing subliminal to consciousness. When the stimulus-mask gap was prolonged from 16 to 100 ms and the stimulus became easier to detect, a fronto-parietal-temporal network became active from about 270 ms after stimulus onset. This activity increased rapidly and nonlinearly as a function of the stimulus-mask delay, possibly reflecting an "ignition" of information broadcasting across the network. This is a striking result, even though it remains uncertain whether the ignition correlates to (phenomenal) perception or to verbal reporting—which, according to some researchers, are inseparable processes (DeHaene et al., 2006; Cohen & Dennett, 2011; but see Block, 2005).

7.7 How Far Must Recurrent Feedback Reach Down?

If we accept that recurrent activity is required at some level of the visual hierarchy to explain the disruptive effect of backward masking, how far down the hierarchy should this feedback reach? This is where opinions in the field start to diverge dramatically. One school of thought, vocally supported by Rees and colleagues (Rees et al., 2002), states that neural activity in V1 is necessary for visual awareness to arise but by and large functions as a "relay station" because the type of information coded by V1 neurons does not match the visual information we become aware of. In this view, awareness is associated specifically with networks higher up the hierarchy, involving fronto-parietal-temporal networks important for coding and interpreting objects and spatial scenes. Another school maintains that feedback from higher to lower areas, including V1, is key to awareness, whereas frontoparietal networks primarily subserve behavioral reporting (Lamme & Roelfsema, 2000; Pascual-Leone & Walsh, 2001; Bullier, 2001a,b; Silvanto et al., 2005; Block, 2005).

Proponents of the "higher-areas-only" view argue that the firing behavior of V1 cells strongly differs from the psychophysics of conscious vision. When we blink our eyes, for instance, neural responses to photic stimulation of our retina, LGN, and V1 cease temporarily, yet we do not experience an interrupt in perception ("blink suppression")—pointing to an integrative process operating beyond V1 that ignores or bridges the temporal gap. Secondly, we cannot subjectively identify the eye to which a stimulus is presented (a task called utrocular discrimination), but many V1 neurons differ strongly in their sensitivity to inputs from the left or right eye. In binocular rivalry, each of two images reaching one of the eyes will activate a subset of V1 neurons, tuned to specific stimulus features (Sheinberg & Logothetis, 1997). Subjectively, however, we do not experience a merged view of the two images but rather a spontaneous alternation between one dominant image and the other, in an all-or-none manner.

In contrast to V1, neurons in extrastriate area V4 show clearer patterns of suppressed or enhanced spiking activity consistent with perceptual rivalry (Leopold & Logothetis, 1996).

In human fMRI experiments, Tong et al. (1998) brought face stimuli into rivalry with images of buildings. They observed that BOLD signals from the fusiform face area correlated with face perception, whereas signals from brain areas analyzing complex spatial layouts (e.g., the parahippocampal place area) correlated with seeing buildings. Another argument to situate neural correlates of visual awareness at higher cortical levels comes from studies on the “attentional blink.” This refers to the frequent failure of subjects to detect a target stimulus when this is presented in between 200 and 500 ms after a first, attention-grabbing stimulus was presented. Normally, presentation of a visible word is accompanied by large ERP activity over the occipitotemporal cortex. This activity patterns is intact up to about 180 ms after a word has been presented that remains invisible in the blink period, suggesting that such early occipitotemporal activity does not correlate with conscious vision (Sergent et al., 2005).

Strong as these arguments may appear, they do remain *correlative*, meaning that they rely on a lack of correlations between certain measures of V1 activity and subjective experience, without demonstrating a causal connection. The arguments derived from single-cell activity in V1 can be countered by proposing a dependence of conscious vision on specific configurations of *population activity* in which single-cell sensitivity to blinking, the monocular nature of input, and other factors become less irrelevant. In support of this, BOLD signals in V1 have been reported to reproducibly fluctuate with transitions in conscious perception under binocular rivalry conditions (Polonsky et al., 2000). To conclude that neural activity in V1 is not causally necessary for conscious percepts because it is not expressed in the right “format” amounts to a non sequitur. This is because perception may directly and causally depend on V1 activity even though the properties of single V1 neurons do not directly explain why we experience percepts the way we do. As I will argue later, perception may be situated at a different organizational level of brain activity than is manifest from V1 single cell firing, and thus its global properties may differ radically from the low-level elements that contribute to it. In the spirit of Gilbert Ryle (1949), a visitor to a university commits a category mistake when claiming that the bricks of a building could never be part of the university because simple stones can never explain its distinguishing feature, intellectual performance.

The other school of thought, adhering to the concept of “recurrency all the way,” also relies on several mainstays. The first one is the electrophysiological evidence for a correlation between recurrent feedback and visual perception (Bullier, 2001b; Pollen, 2011). For instance, Supèr et al. (2001b) trained monkeys to tease apart a textured figure displayed against a differently textured background while recording multiunit spike activity in V1. When the monkey made a behavioral response, expressing it had seen the figure, the spike rate in V1 assumed a higher level than when the monkey had not seen the figure. This “seen–unseen” difference in V1 activity only became visible after an initial response peak ended at about 100 ms after stimulus onset, and lasted for as long as the figure was presented. This delayed appearance of a seen–unseen difference was interpreted as being

caused by recurrent feedback from higher areas. A second argument relies on causal interference with neural activity in V1 and higher cortical areas. Using carefully timed TMS pulses on area V1 and MT/V5, Walsh and colleagues obtained a series of results indicating that V1 activity is required for visual awareness (Pascual-Leone & Walsh, 2001; Silvanto et al., 2005).

At this moment, I see no conclusive evidence to either accept or reject the “recurrency all the way” view, and the jury is still out. This is because the supportive electrophysiological evidence is correlative, and the TMS technique is not sufficiently specific to rule out other explanations. What we need are more conclusive experiments conducting highly specific, causal interventions at different levels of the visual hierarchy. One of the arguments against V1 and recurrent feedback being necessary for conscious vision is provided by several patients with confirmed V1 lesions, who were aware of moving visual stimuli presented in their blind field (ffytche & Zeki, 2011). However, it is hard to exclude that some parts of V1, or adjacent parts that could substitute for V1 processing, may have been spared. Finally, we should not forget to consider *why* recurrent activity would be sufficient or necessary to generate conscious percepts—a problem returned to in chapter 9.

7.8 Oscillations in Conscious and Nonconscious Processing of Sensory Inputs

In addition to slow oscillations, the brain displays a wealth of other rhythmic patterns, and different neural mechanisms are at work in generating them (Buzsaki & Draguhn, 2004). Brain oscillations may arise from a simple circuit of neurons incorporating some type of inhibitory feedback, so that an excitatory effect from a cell *A* onto cell *B* results in a subsequent inhibition of cell *A* by cell *B* (see figure 7.9A). Even the membrane of a single neuron can already incorporate the principle of electric feedback to generate an oscillation. For instance, a given ion channel may open up to depolarize the cell, and this depolarization triggers the opening of another type of channel that repolarizes the membrane back toward rest. As GABA interneurons are ubiquitously found across brain structures and form inhibitory feedback circuits with principal cells, oscillations are a wholesale business for the brain. However, is there any type of oscillation specifically dedicated to conscious as opposed to nonconscious processing? Besides delta oscillations, alpha (8–13 Hz), beta (14–30 Hz), and gamma activity (30–100 Hz) have been examined, mainly so in the visual system.

Gamma activity attracted a great deal of attention when Gray and Singer (1989; Gray et al., 1989) linked these oscillations to integrative visual processing in V1, but gamma oscillations as such had been observed earlier on in V1 as well as in other sensory systems (e.g., Eckhorn et al., 1988; Di Prisco & Freeman, 1985; Boeijinga & Lopes da Silva, 1988). The excitement was boosted later by the discovery of gamma activity during wakefulness and dreaming in humans, as revealed by magnetoencephalography (MEG; Ribary et al., 1991; Llinás & Ribary, 1993).

A great deal of the hustle and bustle was due to the conjecture that gamma oscillations might help solve the binding problem (see chapter 3). Potentially, these oscillations could provide a vehicle for neurons to synchronize their firing when coding feature elements that ought to be bound into coherently perceived objects (Von der Malsburg, 1981, 1995, 1999; cf. Milner, 1974). When Gray and Singer presented single moving bars in the visual field of anesthetized cats, they recorded gamma activity in the 40 Hz range in the LFP of primary visual cortex. At a finer scale, firing activity jointly recorded from local neurons (“multi-unit activity”) also turned out to fire in a gamma rhythm and did so at specific phases of the LFP gamma cycle (see figure 7.9B). This phase-specific firing is also called phase locking. Strikingly, gamma oscillations are coherent across spatially segregated cortical columns (Gray et al., 1989). According to Gestalt principles, common motion of spatially separate stimuli can be used as a perceptual binding cue—for instance, when we identify a deer running away from us in a forest, partially occluded by trees standing in front. When Singer and colleagues presented light bars in the visual field of anesthetized cats, they found that it made a big difference to the synchrony among cells in different orientation columns of the visual cortex whether a single bar or two bars with different orientation were presented. As before, cells in different columns but showing overlap in receptive fields showed synchronized gamma frequency firing, but this correlation was abolished when the same cells were stimulated by light bars of different orientation (Engel et al., 1991). This result confirmed that temporal correlation may be used to group cells representing one and the same object together.

Despite these encouraging results, later studies raised strong arguments against the “binding by gamma synchrony” hypothesis. Experimentally, neural synchrony was missing under conditions where perceptual binding was required, such as discrimination of figure-ground or object continuity (Lamme & Spekreijse, 1998; Roelfsema et al., 2004; Craft et al., 2007; cf. Shadlen & Movshon, 1999). As already shown by the Singer group, gamma oscillations in sensory cortices are found under anesthesia and thus do not selectively occur in the conscious state. They strongly depend on the nature of the visual stimulation applied; single moving gratings elicit much stronger oscillations than plaid stimuli, made up by overlapping gratings, and perceptual coherence does not correlate to plaid-induced gamma synchrony (Lima et al., 2010). Moreover, gamma-range LFP oscillations and synchronous gamma-band spiking have also been described in brain areas implicated in nonconscious processing, such as a respiratory center in the brain stem, the ventral striatum, and the cerebellum (Sebe & Berger, 2008; Van der Meer & Redish, 2009; Kalenscher et al., 2010; Person & Raman, 2012; Bosman et al., 2014), indicating they have a more general role than subserving conscious processing.

Besides these experiments, theoretical arguments have been raised against binding-by-synchrony. Shadlen and Movshon (1999) argued that synchronized firing of neurons that represent features of the same object presupposes that the binding problem has been solved already beforehand. Neural mechanisms taking care of the desired synchronization must

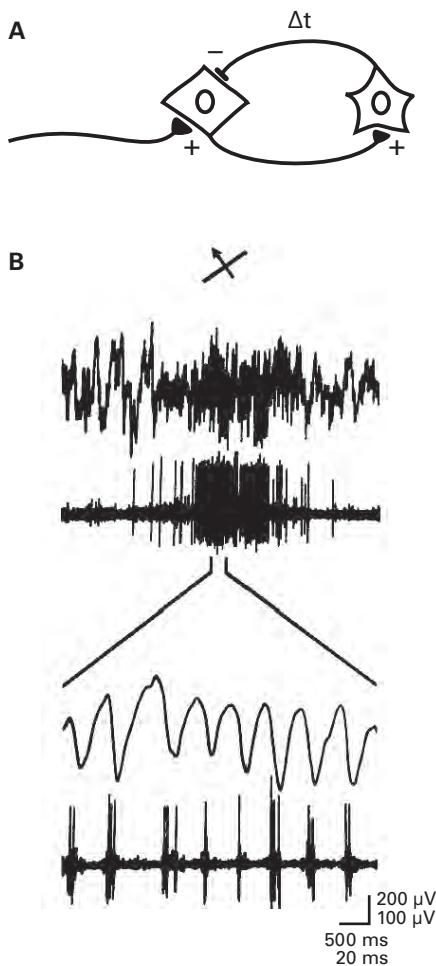


Figure 7.9

Oscillations and gamma rhythmicity in primary visual cortex. (A) Simple circuit diagram consisting of two neurons capable of generating an oscillatory pattern of activity. Triangular and flat synapses are excitatory and inhibitory, respectively. If the left-hand neuron receives an extrinsic excitatory input and begins to fire, it will cause subsequent spiking in the right-hand neuron. If the left-hand neuron is inhibited by the right-hand neuron via an appropriate time delay, its firing rate will temporarily decrease. The next excitation–inhibition cycle will begin once the inhibitory input decreases and sufficient extrinsic excitation is present. At represents a time delay. (B) Presentation of a moving light bar through the visual field elicits gamma rhythmicity in primary visual cortex (area 17) of an adult, anesthetized cat. Top panel shows an optimally oriented bar stimulus moving in the preferred direction of the cells locally recorded. Upper trace (slow time scale): stimulus onset elicits high-frequency activity in the local field potential. The trace below it has been filtered to reveal multiunit activity (i.e., unsorted spiking activity from many neurons close to the recording electrode). The lower two traces (fast time scale) show rhythmic oscillations in the local field potential (35–45 Hz) and multiunit firing correlated in phase with the negative peaks in the local field potential. (B) from Gray and Singer (1989).

be somehow informed beforehand about which neurons should be synchronized and which not. This prior information can only be derived from some other mechanism, ultimately informed by sensory inputs, specifying which features belong together and which do not.

Finally, modulation of gamma-band activity has been found in relation to a host of cognitive processes, such as working memory (Tallon-Baudry et al., 1998), attention (Fries et al., 2001), memory encoding and recall (Colgin et al., 2009), reward anticipation (Van der Meer & Redish, 2009; Kalenscher et al., 2010), and behavioral inhibition of impending motor responses (Van Wingerden et al., 2010). Returning to sensory processing, recent optogenetic experiments by Cardin et al. (2009) suggested a role for gamma oscillations in the inhibitory gating of somatosensory responses to whisker movement in mice.

If gamma activity is not specifically dedicated to conscious processing, what could be its general function? One obvious possibility—also applicable to other types of oscillation—is to prevent hyperexcitability or runaway excitation, as neurons rarely fire during the inhibitory phase of the gamma cycle. Secondly, during the excitable phase of an oscillatory cycle, many neurons in the area generating LFP with gamma power⁶ will fire nearly in sync. If these neurons emit excitatory projections to the same neurons in a target area, the EPSPs elicited in the postsynaptic cells will tend to add up (temporal summation; see figure 2.4), so their net effect on the target cells is stronger than when neurons fire asynchronously (Salinas & Sejnowski, 2000). Reliable postsynaptic firing following strong presynaptic activity may also boost spike-timing dependent plasticity (Abbott & Nelson, 2000). Finally, gamma rhythmicity may regulate the timing and intensity of information trafficking between brain structures (Womelsdorf et al., 2007).

As compared to gamma activity, less is known about the function of *beta-band oscillations* (14–30 Hz) in cognitive processing. Beta oscillations can be recorded over large expanses of neocortex, including regions for sensorimotor processing (supplementary and premotor areas; sensorimotor cortex; Murthy & Fetz, 1996a,b; Hari & Salmelin, 1997; Salenius et al., 1997; Cheyne et al., 2003) and primary sensory and higher associational structures such as visual cortex, piriform cortex, and entorhinal and parietal areas (Kay & Freeman, 1998; Boeijinga & Lopes da Silva, 1988; Buschman & Miller, 2007; Siegel et al., 2008). When field potentials were recorded within sensorimotor cortex, beta oscillations were associated especially with activity of deep layer (L5) pyramidal neurons (Murthy & Fetz, 1996a; Baker et al., 1997), whereas gamma activity is more prominent superficially (Smith et al., 2013).

It is still unclear whether beta oscillations subserve one common, general function or a range of distinct functions; they also vary considerably per brain area and occur in distinct frequency bands within the 14–30 Hz range. Early studies suggested diverse, modality-independent roles, such as enhanced expectation and attentiveness (Lopes da Silva, 1991; Bouyer et al., 1992; Kay & Freeman, 1998), working memory (Tallon-Baudry et al., 2001; Siegel et al., 2009), or planning and execution of movement (Hari & Salmelin, 1997; Baker et al., 1997). Could there be any unified function for beta-band activity? Recently, Engel

and Fries (2010) proposed that beta-band activity, regardless of the brain area where it is found, subserves the maintenance of the current sensorimotor or cognitive state or “set” the organism is in.⁷ When an animal expects a continuation of the current state of the environment and its own body, and this expectation takes priority over potential external stimuli, beta oscillations are strong, whereas a violation of expected sensorimotor patterns is coupled to diminished beta power and increased gamma activity. In this scenario, beta oscillations maintain a stable prediction of, and attentiveness to, upcoming sensory states by reinforcing communication from higher, associational areas to primary cortices in a top-down fashion (cf. Buschman & Miller, 2007). In contrast, bottom-up communication in the visual cortical system is characterized by gamma-band activity (Bosman et al., 2012).

At present there is no convincing evidence for a privileged or selective coupling between beta-band activity and consciousness. Beta oscillations are clearly manifested in brain structures that can be missed when it comes to basic conscious sensation—such as (pre-) motor cortical areas, hippocampus, and striatum (Kühn, Kempf, et al., 2008; Moroni et al., 2007)—which argues against their mere appearance’s correlating with conscious representation. This is confirmed by EEG studies in epilepsy patients who became unconscious during generalized tonic–clonic seizures. Similar to gamma activity, beta-band phase synchrony between intracranial electrodes was more widespread in the unconscious than conscious state (Pockett & Holmes, 2009). This reinforces the point that coherent and widespread high-frequency oscillations can be “bad” for sustaining consciousness. Whereas in general beta and gamma oscillations are not uniquely associated with conscious representations, their specific occurrence in sensory thalamocortical systems may still play a role in their genesis, especially at the assembly level. If the respective coupling of beta and gamma oscillations to top-down and bottom-up visual processing is taken as leading evidence, it is straightforward to hypothesize that beta activity may reflect predictions of future sensory states derived from an internal “model” (cf. Lee & Mumford, 2003; Rao & Ballard, 1999; Bastos et al., 2012), whereas gamma activity communicates actual sensory evidence that gains salience if it comes unexpectedly.

It is appropriate to close this chapter with the EEG rhythm where it all began—the alpha rhythm. Originally, cortical alpha rhythm (8–13 Hz) was found in the EEG across large parts of the scalp, but particularly in parieto–occipital areas, during relaxation in the awake state and during suppression of task-irrelevant neural activity (Berger, 1929; Lopes da Silva, 1991; Palva & Palva, 2007; Buzsaki, 2006). For a long time, alpha activity was regarded as an “idling” rhythm occurring in resting or idling modes, or as a mechanism for cortical inhibition (Adrian & Matthews, 1934; Klimesch, 1996; Pfurtscheller, 2001). Indeed, anesthesia induces an increase in alpha-rhythmic activity, particularly over frontal areas (Supp et al., 2011).

Recent studies that took into account the phasing of alpha oscillations, however, do suggest a link between alpha activity and whether we will consciously perceive a stimulus or not. In the visual system, the stage was set by scalp EEG and MEG recordings from

healthy humans performing a visual detection task. The prestimulus power of alpha oscillations over posterior cortex predicted whether visual stimuli would be perceived or not (Ergenoglu et al., 2004) or whether gray levels of stimuli were perceived as distinct (Van Dijk et al., 2008). When alpha power increased, the likelihood of detection or discrimination decreased, in line with an inhibitory function of alpha activity in perception (but see Linkenkaer-Hansen et al., 2004, for different results on somatosensation). Next, Palva et al. (2005) discovered that ongoing alpha-band oscillations in human MEG recordings phase lock to somatosensory stimuli that become consciously perceived, whereas this effect was not significant for nonperceived stimuli. This means that the timing of alpha oscillations relative to stimulus onset—as specified by their phase—is highly consistent when the stimulus will be perceived.

In line with these findings, Mathewson et al. (2009; see figure 7.10, plate 8) found that also in the human visual system, gauged by EEG, the phase of the alpha cycle, relative to visual stimulus onset, matters for conscious detection in a masking paradigm. When the target was presented at the same time as an alpha wave reached its trough, subjects had more difficulty detecting it. During the trough phase, neocortical activity is likely suppressed at about 100 ms following target onset. When presented outside the trough, the stimulus was more likely to be consciously perceived. These results refine the “inhibition” hypothesis of alpha activity in that the effect is not equal across the alpha cycle but takes the form of a “pulsed inhibition,” expressed during only part of the cycle.

Yet, these studies do not reveal whether alpha oscillations are causally relevant to subsequent perception or merely correlate with it, leaving open the option that some other, unknown process regulates perceptibility. Confirming a causal role, Romei et al. (2010) found that trains of TMS pulses, rhythmically applied at alpha frequencies, modulated target visibility, whereas this effect was not found in theta or beta bands. Imposed alpha rhythmicity hindered detection of stimuli presented in the visual field opposite to the stimulated hemisphere but facilitated detection in the hemisphere lying on the same side as the visual hemifield in which the visual target was shown.

Why would conscious representations be supported in any way by alpha oscillations? Intuitively, we are inclined to think of consciousness as a continuous “stream” of experience (James, 1890), but this subjective continuity need not imply a temporal continuity in the neural substrate. Possibly, the excitatory phase of alpha oscillations packages or “frames” the information in a brief slice of time, and a subsequent cycle represents an updated snapshot relative to the previous one (cf. Gho & Varela, 1988; VanRullen & Koch, 2003). The inhibitory phase of the cycle may then be a “blackout” moment during which information can be rapidly trafficked from low to higher areas and vice versa, enabling the system to assemble a representational update. Gho and Varela (1988) presented two brief light flashes in close succession and showed that the alpha cycle phase at which they are presented matters for subjects perceiving the flashes as being simultaneous or occurring sequentially. An important route for testing these enticing ideas will be to examine how

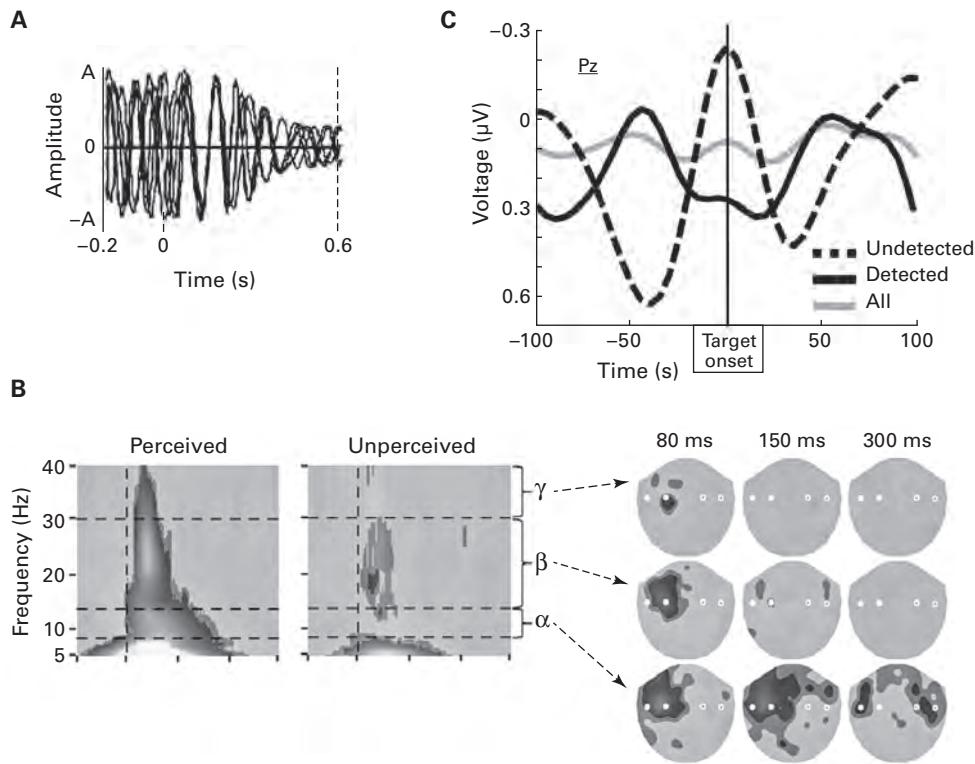


Figure 7.10 (plate 8)

Phasing of alpha oscillations relative to perceived and unperceived stimuli. (A) Simulation of magnetoencephalographic (MEG) traces of oscillations recorded from cerebral cortex. Superimposed traces show 8–12 Hz oscillations plotted relative to onset of a weak electrical stimulus (0.2 ms) applied to the tip of an index finger. Note how ongoing alpha-band oscillations transiently phase lock to stimulus onset, that is, the oscillation has a consistent phase relationship with stimulus onset. (B) Time–frequency plots of phase locking strength over primary somatosensory cortex (contralateral to stimuli). Color transparency scales with the statistical significance of phase locking. Only consciously perceived stimuli are associated with phase locking in alpha-band oscillations. The right-hand side shows the topographic scalp distribution, across three different time points relative to stimulus onset, with predominant alpha-band phase locking across frontoparietal areas. (C) Averaged event-related potentials recorded from the midline at the level of parietal cortex (channel Pz) in a visual target detection task with metacontrast masking. Around target onset, undetected targets (dashed curve) are associated with a phase that is opposite to that of detected targets (solid curve). When detected and undetected targets are taken together, phase locking is absent (gray curve). This indicates that stimulus timing relative to the excitatory and inhibitory phases of the alpha cycle is important for conscious perception. (A, B) from Palva et al. (2005), with permission from Elsevier. (C) from Mathewson et al. (2009), with permission from the *Journal of Neuroscience*.

perception and EEG waves are timed with respect to spiking activity in visual and higher cortical areas and manipulate activity of regionally confined cell populations during selective phases of the cycle.

In summary, the neurophysiology of sleep, anesthesia, and conscious states paints a very diverse and sometimes opaque picture of brain dynamics associated with conscious processing. Even though there are many loose threads to be followed, we can begin to discern some important motifs. Comparing sleep stages with wakeful states, we see evidence emerging for the primary role of sufficiently prolonged Up states in the neocortex, featuring high levels of depolarization, irregular firing, and a desynchronized EEG signature. Also pathophysiological and computational studies emphasize the importance of having sufficiently decorrelated activity rather than global synchrony, which is nonetheless compatible with coherent micro- and transregional oscillations in conscious states. The main lesson drawn from studies on replay is that this form of coordinated, multi-area activity is unlikely to correspond to conscious processing, challenging any simple implementation of “neural coalitions” embodying consciousness. In line with this, findings on anesthesia highlight the prominence of sustaining long-range, extensive rather than local, restricted interactions between distributed neural assemblies, especially in the posterior neocortex. Recurrent interactions from higher to lower cortical levels are probably important in perceptual microgenesis, yet more work is needed to find out which components of top-down connectivity are essential for perception or behavioral reporting. Whereas in this system gamma and beta oscillations may exert locally regulating and anticipatory, set-maintaining functions, respectively, we are witnessing the alpha oscillation returning to the forefront of the field, with an emerging role of its phasing in perceptibility.

8

Requirements for Conscious Representations

In the act of diving into shallow water a young soldier experienced a sudden pain in the neck and at once lost the power of his limbs. He was pulled out of the water by his comrades and soon regained some power in the left upper and lower limbs. [...] On the left side of the body profound sensory changes of an unusual character were found on the trunk and lower limb [...]. The patient replied to all forms of stimuli, but the sensations evoked by the different forms of stimulation were all described as similar. Tactile, pressure, thermal and pain stimuli were not recognized as such, but were described as “queer” or “funny,” “like electricity” or as “making him laugh.”

—R. M. Stewart (1925)

8.1 Integration

Up to this point, we have particularly highlighted what the problems are when trying to build conscious representations using neurons. It is much less clear how to solve these problems, at least conceptually. Phenomenal consciousness remains the ultimate enigma in mind–brain science as it has been in the past, and no ready-made solutions can be jotted down. What we can do is first begin to outline the requirements that a brain system needs to fulfill to generate a conscious representation (this chapter), and then get a handle on the neural mechanisms realizing these requirements (in chapter 9).

Most likely, the brain remains conscious when “stripped” from its capacities for emotion, episodic memory, and motor functions. What would be the bare-bones necessities for a “minimal” brain system sustaining conscious sensations? Classic neural network models are insufficient to provide such a basis, among other reasons because they fail to explain why it is that we experience some sensory inputs as visual and others as olfactory. A related problem entails that neural-network-like structures abound elsewhere in nature, even in inorganic form, and this problem of panpsychism resurfaces when we consider accounts based on information theory. If we probe the demands that a conscious system actually poses and try to translate these demands into plausible neural mechanisms, will we then still face the same panpsychic trap, or will we arrive at a brain system sufficiently specific to be distinct from inanimate systems?

Viewing the problem of consciousness from a bottom-up anatomical and physiological perspective, we saw that an appropriate brain *infrastructure* is not enough to sustain consciousness. Thalamocortical and associated systems not only should be wired up correctly but also should assume an appropriate *physiological state* to do the job. Job fitness is all too easily lost when we fall into the oblivion of deep sleep or inhale an anesthetic gas. In the case of anesthesia, *local cortical processing* retains an uncanny similarity to conscious states. So why is it then—in the face of such local processing capacities—that the entire system fails to meet the requirements for conscious processing?

Let us first go back to the rock-bottom biological function of consciousness. Organisms roaming about in an environment full of threats and opportunities for survival and reproduction face the representational problem of needing rapid updates about the world around them and about their own states (see chapter 6). They need a quick-and-dirty representation of the objects and events occurring in their immediate range of action. This situation sketch should be rich and complete enough to capture events and object properties in all relevant modalities and to have actions conducted in a context-sensitive manner. A first requirement on a brain system expressing this capacity, as already noted, is its ability to *integrate* information from separate sensory sources into a single, multimodal representation of the organism's situation. A fox chasing a hare fleeing through the grass will benefit from integrating visual, olfactory, and auditory information, as this integration provides more accurate estimates about the hare's whereabouts and predator-evading tactics than either sense alone. Different features of an object, such as the eyes, nose, and mouth of a human face, can be glanced almost simultaneously, in a global facial expression, but may also be scrutinized separately, by focused spatial attention. Bits and pieces of information are gathered and integrated across short stretches of time, such as when we merge the sight and sound of somebody calling us from a distance into a unified percept, even though the sound arrives slightly later at our ears than the visual input hitting the retina.

We have been coining this notion of *integration* before, but what does it actually mean? Here I use it to denote how information derived from different sensory sources is merged into a *single*, immersive, spatiotemporally ordered experience, instead of various disparate experiences. Our brains merge the two slightly different photic patterns impinging on our left and right eye into a single visual representation enriched with the quality of depth, instead of two separate, nonstereoscopic views. Another example is the merging of information from two or more modalities: we perceive a ventriloquist's voice as coming out of the puppet's mouth, skillfully moving in sync. We can close our eyes and focus attention on the ventriloquist's voice alone, but the point is that multimodal integration happens naturally: we cannot help doing so, whereas in theory there would be an option to process sight and sound separately.

Also within a single modality, our brains construct objects and events by integrating information from different submodalities. To an antelope, a cheetah sneaking through distant clumps of grass will present itself as a single object, despite its being composed

of color, shape, texture, and motion features processed in different nodes of the visual system. The more general term *integration* can be subdivided in several processes, including the *binding* of different sensory features into a coherently perceived object. Classically, the binding problem is concerned with the integration of different *kinds* of features in a common object (e.g., its color and shape; see chapters 3 and 7). In the case of a cheetah moving through grass, we are also dealing with the perceptual *grouping* of features of the same kind moving coherently (e.g., patches of coat; this involves the Gestalt laws of grouping, e.g., Köhler, 1940). That integration is a real process, fundamental to our perception, is vividly illustrated by its breakdown in the case of brain lesion: in visual agnosia we saw how patients can still perceive line elements but fail to perceive the whole object (see chapter 3).

Sometimes it has been claimed also that differently localized elements of a spatial situation are integrated. On the one hand, the elements are part of one and the same experienced situation and can thus be said to be integrated; on the other hand, this use of “integration” is somewhat confusing because the elements are often detected across saccades as they are spatially *segregated*, and so we also have to look for neural mechanisms ensuring ways to separate (rather than integrate) streams of information belonging to different objects. We need ways to sort and situate streams of information, setting apart objects from each other.

The functional-perceptual notion of “integration” I am using here should be kept apart from *neural measures* attempting to describe integration of information—or communication—between brain areas, such as measures of EEG or spike train coherence (Tononi & Edelman, 1998; Tononi, 2004; Seth et al., 2006; Barnett et al., 2011). Finally, the two connotations of “integration” used here (single experience and binding–grouping vs. segregation) are not exhaustive; below I will elaborate on the emergence of phenomenal content or meaning.

8.2 First-Person Perspective, Time and Space

Even if our brains fulfill the requirements for integration, this does not guarantee we will perceive the world and our own body from a biologically appropriate—as opposed to arbitrary—viewpoint. Normally we see the world in a first-person perspective that dominates the spatial framework of our experiences (cf. Searle, 1992; Churchland, 1995; Metzinger, 2000) and is tightly coupled to our head position and orientation. As our heads rotate and our eyes move in saccades, visual inputs continuously change over time, yet our brains signal that we stay where we are and the environmental situation is invariant. The first-person perspective extends to other senses beyond vision, and even to having a “temporal perspective,” the sense that we are directly experiencing the present and do not live in the past—in our retrieved memories—or future.

This temporal aspect of experience involves multiple memory systems operating across partially overlapping time domains, such as iconic memory,¹ working memory, and

long-term memory (cf. Husserl, 1928; Allan, 1979; Fuster, 2003; chapter 3). Especially working memory, actively holding recent information in an online buffer and using this information prospectively to negotiate future events and actions, is probably crucial for the continuity of subjective experience, for the embedding of current events in the context of the past and future. Nonetheless, patients with severely impoverished working memory do have conscious experiences and recollections from long-term memory (see chapter 3), which raises the question of whether a sense of time—as empowered by short-term memory—is actually an *a priori* attribute of consciousness as was argued by Kant (1787) and others. Patient studies lend credibility to a scenario in which consciousness for verbal or visuospatial items can exist on a near-instantaneous basis without acutely relying on working memory holding such items (Hanley et al., 1991; Vallar, Papagno, & Baddeley, 1991).

Returning to spatial aspects of conscious experience, our “egocentric” view of the world may seem straightforward. However, the apparent unification of sensory inputs becomes less obvious when we realize that different sensory and motor modalities and submodalities may use different spatial frameworks and reference points for representation. Knowledge of “where I am in space” is not naturally set in an egocentric, visually dominated framework, because our view of the environment does not provide unambiguous information about where we are. When we stand in a room with four visually identical windows and corners, our view of any particular window or corner does not tell which of the four windows or corners we are facing. Information on previously traversed paths and body rotations in space is needed to know where our body is situated in space (Tolman, 1948; O’Keefe & Nadel, 1978; Worsley et al., 2001; McNaughton et al., 2006).

When one is designing a robot with a capacity to navigate through unknown territory, the robot’s performance will benefit from having an allocentric framework in addition to an “egocentric” spatial view. An allocentric representation of space does not depend on one’s own immediate viewing direction but maps the organism’s position onto a coordinate system independent from this, taking into account its distance to objects that are not immediately perceived but have been encountered in the past (cf. O’Keefe & Nadel, 1978; Mittelstaedt & Mittelstaedt, 2001; McNaughton et al., 2006). When you are sitting on the couch in your house at night, you can probably easily point your finger in the direction of where your car or bicycle is located, even if these are not visible. Our daily-life first-person, egocentric perspective is less trivial than it seems. Why not view your environment, with you in it, “from above,” so to speak, and see your own body navigate through your house like a Pac-Man? Note the paradox buried within this question, as our imagination allows us to “jump” to a different perspective but also remains bound to directed viewing.

There are further reasons to scrutinize how solid the apparent unity of the first-person perspective is. When you rotate your head with eyes open, your experience will be easily dominated by the change in environmental view, but the constancy of your body axis position in space is part of the same experience, albeit perhaps at the fringe of

awareness. Proprioceptors and mechanoreceptors in the skin fail to indicate a change in body posture, and a head rotation elicits a vestibular response matching the shift in visual view, confirming that no net change in body axis position took place. Thus, a change in “egocentric view” is not necessarily accompanied by an experienced change in one’s own body position, raising the question of what the “real” egocentric framework is. The plausible answer is that there is no single, valid egocentric framework. Although the visual and somatosensory frameworks share an “egocentricity” in that each modality alone does not endow us with an objective, allocentric viewpoint, it has become clearer by now that different sensations are primarily anchored either to individual sense organs, such as the eyes, or to combined sensory inputs resulting in an inferred physical attribute, such as own-body position. We will come back to this in chapter 9.

That the “unity” of our first-person perspective is not self-evident is also highlighted by occasions when frameworks anchored to different modalities become misaligned. Most likely you will remember the sensation aroused when you were a kid, standing upright and starting to spin around your body axis, arms stretched out widely. A sudden stop would make you giddy, with your brain all confused, trying to reconcile the constancy of the visual input with the ongoing directional flow in the semicircular canals of your inner ear. The visual world view will be experienced as dominant, but vestibular inputs are too powerful to be ignored in this situation.

Sensory misalignments can be more dramatic than this, especially in neurology. In a particular form of autoscopic experience a patient may be walking through a city, turn around a corner, and stand face to face with ... herself! (See Brugger, 2002.) A patient whose right angular gyrus was electrically stimulated in the course of evaluation for treatment of epilepsy was subject to out-of-body experiences, marked by a visual first-person perspective moving away from one’s own body position. This patient described her experience as follows: “I see myself lying in bed, from above, but I only see my legs and lower trunk” (Blanke et al., 2002, p. 269). Also healthy subjects can be tricked into believing they are somewhere else in space than where their body is physically stationed. Lenggenhager et al. (2007) instructed subjects, standing upright in a familiar room, to wear video goggles and to experience the paintbrush strokes on their back. At variance with normal experience, they looked at the video stream taken by a camera that was positioned *behind* them, projecting the brushstrokes on their own back into their eyes via the goggles. As the subjects felt the brushstrokes on their back but also saw how these movements happened on a body that was visually in front of them, they estimated that their own body position was significantly ahead of themselves, toward the projected body.

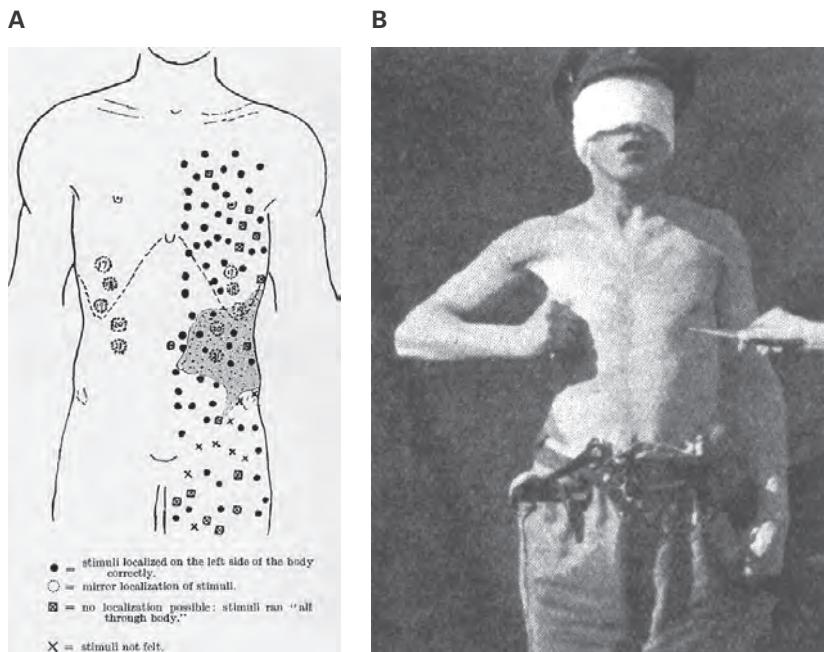
The upshot is not to deny that our experiences are normally set in a first-person perspective but to recognize that this perspective is much more complicated and multifactorial than one might suspect. The actual requirement for conscious experience is not so much to squeeze all sensory information into the same perspectival “format” but to take care that the manifold frameworks used in different modalities remain well-aligned with each other

at consecutive processing stages. That way, we are not continuously tossed around between experiences seen or felt from different perspectives, as one might experience when viewing M. C. Escher's famous 1947 woodcut, *Another World*.

As we already touched upon Kant's notion of time as a necessary attribute of consciousness, it is time to address his concept of space. Could there be experiences completely or largely devoid of spatial aspects? Smell, for instance, is not exactly our most spatially accurate modality. With our eyes closed, a scent of perfume might come from anywhere around us. But leaving the problem of source localization aside, where is it that we smell relative to our body map? We associate odors with the position of our nose—we know we do not smell at the tip of our index finger, for instance. Humans are microsmatic, having low olfactory capacities relative to dogs, but even we have an ability to perceive whether an odor reaches the olfactory epithelium via the nose (orthonasally; associated with an external source) or via the mouth–larynx opening (retronasally; associated with food flavors; Murphy et al., 1977; Small et al., 2005). Spatiality is not alien to modalities different from vision or smell, ranging from taste, touch, pain, and thermoception to less localized body sensations such as visceral sensations, or, more popularly, “gut feeling.” Even in a *Ganzfeld*—in which subjects are facing an evenly lit and contourless visual field—people having both eyes open retain a notion of space, a realization they are looking at a spatially divisible environment and not at a point or 3-D object (Bolanowski & Doty, 1987).

However, it is neurology again that defies classical philosophy. Normally, we find it completely natural that somatosensations are correctly localized; we refer them to spatial positions on our body being stimulated. However, clinically, stroke involving the thalamus, parietal–somatosensory cortex, and adjacent brain structures may result in mislocalization of touch, or even an absence of localization (Halligan et al., 1995; Turton & Butler, 2001). In rare cases, partial spinal cord injury not only leads to a loss of specific sensory quality associated with all sorts of stimuli applied to large areas of skin (e.g., touch, pain, heat, cold; a condition termed “archaesthetic” sensation by Stewart, 1925) but is combined with a complete loss of ability to localize such stimuli. Stewart’s patient with a complex spinal hemianesthesia was unable to point to particular locations on the left side of his trunk when touched and said that the feeling “ran all through him” (p. 309; see figure 8.1).

Returning to space as a suspected a priori condition for vision, we already encountered blindsight, the phenomenon that V1-lesioned patients do not perceive visual stimuli presented within the blind part of their visual field but perform above chance level when forced to guess stimulus features (see chapter 3). Weiskrantz and colleagues found that it is also possible, however, that a patient “feels” or senses that an event occurred without seeing it per se (“blindsight type 2”; Weiskrantz, 1997; Sahraie et al., 2010). Patient D.B., for instance, had his right striate cortex removed and was unable to localize stimuli presented in the left visual field, while still being aware that something was going on. This is the most basic form of consciousness we can discern: a mere *sense of presence*, without knowing what is present, or where it is in space. This is confirmed by the nature of the

**Figure 8.1**

(A) Body surface map of the patient described by R. M. Stewart (1925), suffering from a complex form of hemianesthesia due to spinal cord injury. Hemianesthesia is the loss of somatosensation on one side of the body (which can be partial). Before the patient completely lost the ability to localize stimuli on his left body surface, this map was drawn showing the localization of cold stimuli. At this early stage, tactile stimuli were localized correctly, whereas pain stimuli applied to the shaded area were perceived as occurring on the opposite side of the body. Black circles: at these locations, cold stimuli were correctly localized to the left side of the body. Open dotted circles: stimuli perceived, but at the mirror site on the other side of the body. Squares with crosses: stimuli were felt but could not be localized (they ran “all through body”). Crosses: stimuli not felt. (B) Photograph of the patient’s right-chest-directed motor response to a tactile stimulus applied to a corresponding location on the left side. This mislocalization phenomenon is referred to as “allochiria.” Reproduced with permission, copyright © 2014 BMJ Publishing Group Ltd.

sensations described by blind-born people whose congenital cataract is removed by surgery: “the newly operated patients do not localize their visual impressions ... they see colours much as we smell an odour of peat or varnish ... but without occupying any specific form of extension in a more exactly definable way.”²

8.3 Dynamics and Stability

As psychophysical experiments on microgenesis suggest, the brain needs time to construct conscious representations (see chapter 7). Referring to Helmholtz’s work in the nineteenth century, we may say that groups of neurons gather “perceptual evidence” in favor or against specific sensory inputs, as expressed in changes of firing rate that take time to develop

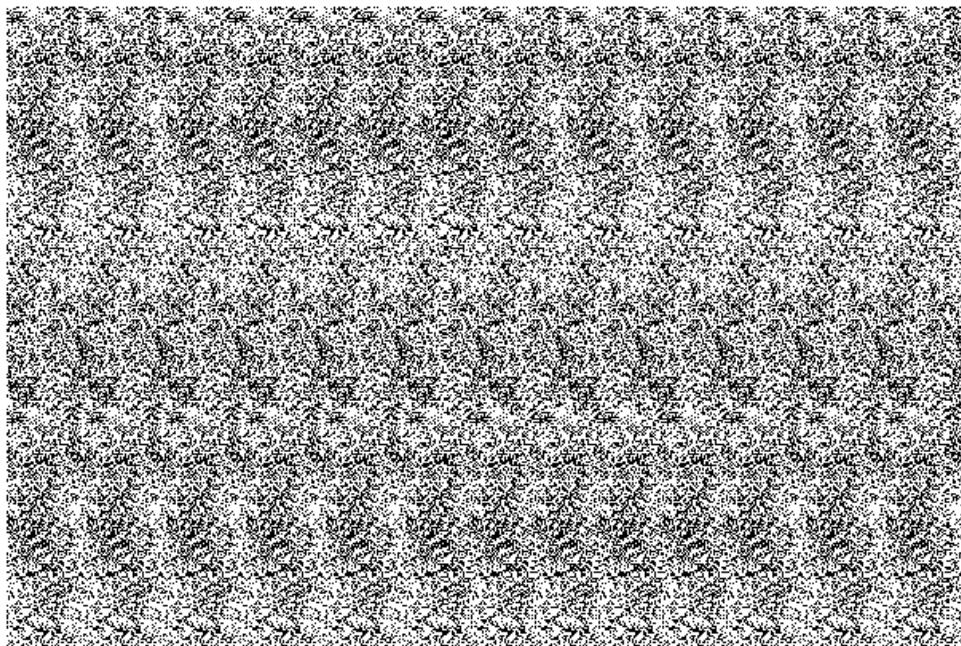


Figure 8.2

Autostereogram as an example of perceptual stability. Popularized by Magic Eye posters in the 1990s, these pictures are based on random dot stereograms designed by Julesz (1971). To see the squares popping out in 3-D, hold the center of the image right up your nose and try to focus on a point at a distance away from the picture. You will see the picture as blurry, but then move it *very* slowly away from your nose until it is about 20–30 cm away. If successful, you will see two squares raised in front of a flat background. You may furthermore see two floating rectangles behind the two boxes. The parallax effect is ascribed to the distance between repeats on the square being ten pixels shorter than on the background. Even if you rotate, translate, or fold the image with respect to your eyes, the depth percept remains stable.

(Cohen & Newsome, 2004; Huk & Shadlen, 2005). Equally as interesting as the *dynamics* of constructional processes is the *quasi-stability* that conscious representations can acquire (Leopold & Logothetis, 1999; O'Brien & Opie, 1999). For example, the Necker cube and duck–rabbit (see figure 4.5) prompt us to switch between two quasi-stable views. Subjects have some voluntary control over switching and can be trained to perceptually switch faster between alternate views, but there is also an upper limit to the speed of switching (van Ee et al., 2005). Thus, conscious representations, at least of objects perceived in a stationary mode, have a certain stability and robustness, ensuring they are not immediately eliminated once the construct is standing.

Figure 8.2 presents a more extreme example of perceptual stability. It may take some practice to see the squares popping out of the image. Once the trick works, it may take quite a bit of effort to “get rid” of this percept. When I first tried, I diverted my eyes toward the periphery of the picture, but the squares kept staring me in the face—a percept stable enough to dominate volitional control. In binocular rivalry, we saw how subjects report

seeing either the left-eye or the right-eye image (see chapter 7). Their perception alternates between the two at time intervals ranging from about 1.5 to 12 seconds, and again they cannot train themselves to switch at arbitrarily high rates (Van Ee et al., 2005).

The main point here is to highlight that neural systems are required to maintain a representation stably, at least for periods of several hundreds of milliseconds up to seconds. In the same vein, systems should not “get stuck” into a fixed representation when salient changes in sensory input command flexibility. If a red double-decker bus passes you on a London street, your visual experience requires a continuous update of its changes in position. “Pasting” a constant velocity parameter into the representation will not do, as changes in speed, regularity, and direction need to be tracked in parallel. In sum, stability is not a structural requirement of conscious representation, but the ability to both sustain stability and allow rapid transitions to novel percepts is such a property.

8.4 Diversity, Interpretation, and Projection

Dominated by visual perception as our thinking about consciousness might be, it is a *multimodal* phenomenon at its very roots. “Multimodal” should be taken broadly, not only including external sensory input but also elements retrieved from declarative and emotional memory, used to construct imagery (Tulving, 1983; Rubin et al., 2003; Daselaar, Porat, et al., 2010; Huijbers et al., 2011). A person struck by blindness remains conscious of all sensory modalities except vision, and can develop exquisite capacities to make finer tactile discriminations than before, as in Braille reading. The modalities and submodalities constitute an enormously rich *diversity* of sensory signals that we somehow need to represent in the quick situational update we call a conscious representation. That the neural system must be able to provide the update *rapidly* is an evolutionary requirement. Unfit for this job would be a brain system reading out sensory inputs from the periphery as long, unlabeled binary strings (see chapter 6) and representing them as such and not in any other way.

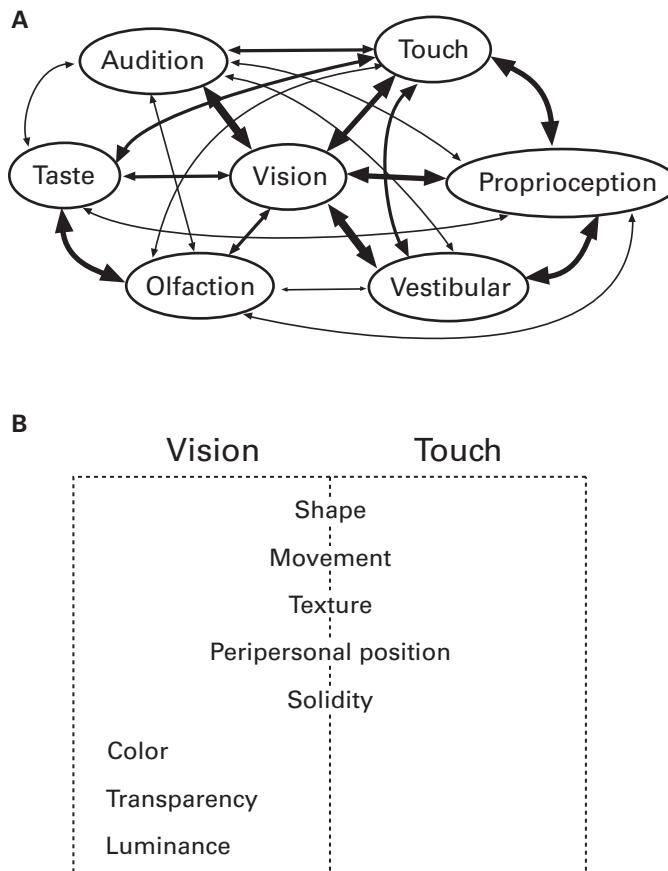
If such a vector-processing system would *only* act to represent sensorimotor information by binary strings but does this as aptly as a normal brain would do, it would instantiate a zombie as it would not sustain experienced *qualia* but instead resort exclusively to *quantita*. If we suppose that a related, yet different system would be conscious of its sensory inputs only as binary strings, it would be at a serious disadvantage because the *experience* of a jambalaya of numbers would severely slow down its speed for making behavioral decisions. Zombies will return in chapter 11, where we will explore the notion that a neural system properly constructing multimodal representations is *bound* to generate qualitative experience.

Unfortunately for those expecting a quick solution, sensory inputs appear to be conveyed from sense organs to brain by and large as unlabeled binary strings (see figure 4.11). No doubt, each sensory fiber is “labeled” in that it originates from a specific type of receptor

that tranduces a particular physical energy change into bioelectric activity, but the problem for a central facility receiving all sensory signals is that it does not have a priori knowledge independently verifying where those signals come from. From the brain's viewpoint, sensory inputs are anonymous. A personal-computer type of solution might be to add a binary "address of sender" code to each sensory message. Besides the fact that a built-in decoder would be needed to identify sensory origin, such an address code would make the system even more cumbersome than it already was. Without using address codes, our brains are able to soak up myriads of sensory inputs, arriving in parallel. Somehow they do offer a fast representational solution by which each modally specific input is not only appropriately ordered in space and time but also rapidly recognizable for what it is—the color blue, a leftward rotation, the buttery taste of a Chardonnay. That this is a general but distinguished ability is underscored by neurology, having described patients who lose the capacity to feel different somatosensory qualities (touch, pain, heat, cold, etc.) when stimulated on parts of their body surface but nevertheless report that something is going on ("funny feeling"; Stewart, 1925, p. 299).

Various approaches can be probed to find a solution. Sensory inputs, processed in the context of other inputs, may become rapidly recognizable by a process of *interpretation*, which includes *classification* as a component. Previously I suggested that the process of interpreting a particular sensory input as having a unique and qualitative nature may be driven by two major factors: the *statistics* of the input (taken as a spatiotemporally varying, multineuron signal—which does not account for qualitative interpretation on its own) and the *relationships* between the input and concurrent representations in other modalities (Pennartz, 2009). The latter account can be exemplified by a subset of our visual experiences being closely related to concurrent tactile inputs—for instance, when we see we put our right hand around a doorknob and concurrently feel the pressure on the hand's glabrous surface. In naturalistic settings, auditory and visual inputs correlate strongly as well. Smell and taste are also tightly bonded to each other but are more loosely associated in time with exteroceptive senses such as audition and vision (see figure 8.3). The shorthand phrasing of the relational concept is "You can only experience what vision is if you also know what nonvisual experiences are." Vision must be compared to, and segregated from, nonvisual percepts. An unambiguous interpretation of a multitude of sensory inputs should be based neurally on the *relational configuration* of input-specific patterns. How such configurations may be coded in the brain will be investigated in chapter 9.

Two different ways of looking at modality identification are, first, to isolate a particular object feature such as the texture of a brick and ask which sensory (sub)modalities you have at your disposal to describe the texture. Vision and touch will easily come to mind. Within vision, different visual attributes such as color, luminosity, specular reflectance, and granularity characterize the stone's surface properties. But you may also hear a subtle whisper as you rub your hand across the brick. By directed hand movements, you effectively include the motor system as a separate modality. Within the brain, brain stem and thalamocortical

**Figure 8.3**

Topology of sensory modalities. (A) Organogram plotting topological relationships between sensory modalities on the basis of estimated, overall correlation strengths. These estimates are subjective and may strongly differ between individuals. Whereas the visual domain is strongly connected to the tactile, proprioceptive, vestibular, and auditory modalities, weaker conjunctions and stronger disjunctions are present between vision, on the one hand, and olfaction and taste, on the other hand. Additional modalities associated with the somatosensory system (thermo- and nociception) are not included in the diagram. Each modality occupies a unique correlational niche in the topology, hence enabling a unique identification. (B) Sensory singularity. Some object properties can be sensed by both the visual and tactile system (shape, movement, texture, position, and solidity), whereas other properties can, in principle, only be perceived by vision (color, transparency, luminance). This is not contradicted by the fact that such perceptual singularities can be strongly associated with sensations in other modalities, such as heat in relation to luminance. From Pennartz (2009), with permission from Elsevier.

motor areas emit divergent projections to many subcortical as well as cortical structures—an output referred to as “efference copy” (Klier & Angelaki, 2008; cf. figure 6.7).

But what do you have at your disposal to identify the color of a cloudless sky in a Monet painting? Can we think of any way of genuinely taking note of color other than by sight? Ridiculous as the question might seem, it does stress that color vision occupies a

unique niche in our perceptual space and is therefore considered a perceptual *singularity* (Pennartz, 2009; see figure 8.3B). No wonder color has been used countless times in philosophy as *the* prime example of “qualia”! Texture, by contrast, occupies a very different niche in perceptual space, positioned at the crossroads between modalities.

A second, related way of looking at modality identification is to ask to what extent information gathered in a given modality X has predictive value for the information gathered via a separate sensing system Y . The more correlated the two inputs in X and Y are, the higher the predictive validity. This view is akin to the theoretical framework of “predictive coding,” developed in the field of vision, holding that what we perceive is not directly our environment but rather the inferred causes of the sensory inputs we receive (Gregory, 1980; Srinivasan et al., 1982; Lee & Mumford, 2003; Rao & Ballard, 1999; Bastos et al., 2012). Under this view, the brain constructs a *model* that is continuously predicting what goes on in the external world to explain the sensory changes reaching our brain, and it uses newly arriving sensory inputs to compute error signals between predicted and actual world states, updating the model.

Compared to our earlier vantage point facing the sheer anonymity of sensory inputs (see chapters 4 and 5), we may now begin to feel less discouraged about the problem of modality identification. If we can conceive a system capable of *self-identifying* modalities, we may even come closer to solving how phenomenal content or meaning may arise. Indeed, the main modalities provide the crudest categories or chunks of phenomenal experience. Sensory inputs arrive in the brain unlabeled, but the sensors do act as *unique signalers*: they *only* become active when a very specific type of input, often tied to a location on the body or its surrounding field, occurs. By this unique signaling, a sensor brings about a unique activity configuration within a neural system (see figure 8.4). Cross-modal interactions have been left out for simplicity here, because the foremost purpose of figure 8.4 is to illustrate how a specific sensory input simply leads to a unique configuration of neural activity across the brain, specifically in the neocortex.

On the one hand, the unique configurations induced by specific sensors help in explaining why many complex, inanimate systems in nature do not qualify for conscious representations. Revisiting our cyclone example (see chapter 5), its water particles are “sensitive” to physical quantities such as temperature and local wind speed, but there is no qualitative difference between particles in being sensitive to some quantities and not others. A local temperature rise is paralleled by an increment in kinetic energy of all particles, big or small, coupled to a local expansion of the cloud. If water particles could be claimed to “signal” anything to fellow particles, following the idea of a larger representational system, they would not do so uniquely. Inanimate systems lack the mechanisms typical for the diversity and richness of multimodal perception.

On the other hand, it is still unclear how an internally differentiated system (see figure 8.3) should represent anything we would associate with *phenomenal content*. Earlier on, we referred to the philosophical-linguistic notion of *intentionality*, holding that mental

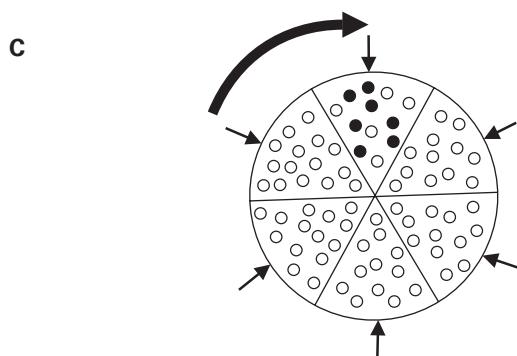
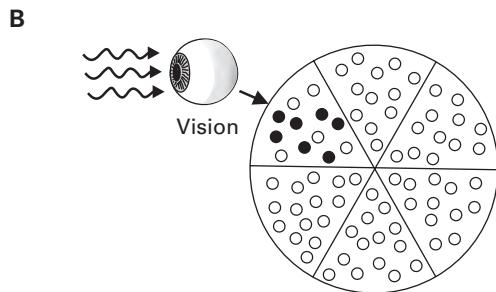
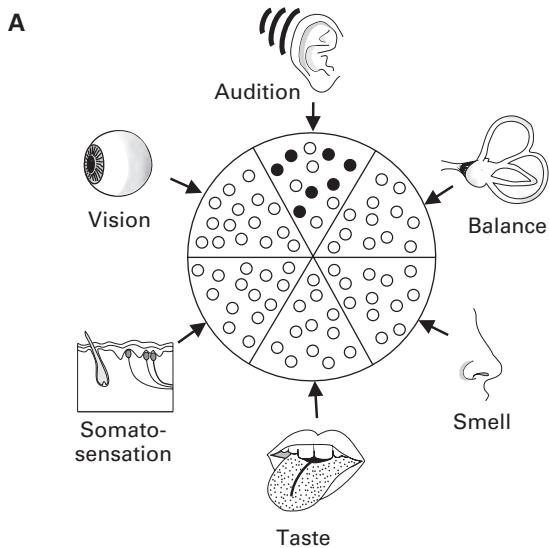


Figure 8.4

Multimodal pie chart comprising six main modalities (audition, balance, taste, smell, somatosensation, and vision; other senses are not included or have been subsumed under somatosensation). (A) Upon auditory stimulation, an ensemble of neurons in the auditory thalamocortical system is activated (black) as a subset of all neurons in this modality. No cross-modal activation is assumed here. (B) Same as (A) but now for vision; a subset of feature detectors in the visual thalamocortical system is activated. (C) Although the scheme of (A) and (B) ensures modality-specific patterns of feature detector activity, the system cannot self-identify the kind of sensory information it is processing at any given time because the pie chart is rotationally symmetrical and the neurons in it have no built-in knowledge of where each sensory input comes from (peripheral sense organs have been removed to illustrate this point). The clockwise arrow underscores that the scheme is rotationally symmetrical with regards to sensory identity, so if all six sensory domains were rotated by 60° and the visual domain ends up at the auditory position, the sensory system lacks information to code this modal change.

states refer to objects, properties, and events in the outside world and thus are *about* something else than themselves (Searle, 2004). If you believe that the ignition of a rocket will result in beautiful fireworks, this mental state is about the rocket exploding in brightly colored sparks, not about the mental (or neural) state doing so.

The current framework acknowledges that conscious systems represent events and objects beyond the system's own events and states. Because the term "intentionality" is often used in a linguistic or cognitivist context (Anscombe, 1965) and carries connotations both to (motor) intentions and higher-order, cognitive belief systems I prefer to avoid, I will refer here to the *constructive* and *projectional* capacity of conscious systems. That is, our brains generate conscious representations having perceptual content, but inherent to this content is that perceived objects—and indeed whole situations or scenes—are projected to, or situated at, different locations than the physical position of our brains. Integral to healthy, full-blown representations is that they are spatially ordered, including the external locations at which objects are situated.

Returning to figure 8.4, a problem yet unsolved is that the pieces in the pie configuration are all interchangeable. The pie is rotationally symmetrical and offers no built-in mechanism to say which portion is concerned with audition and which with vision (except if the afferent sources of each piece are known; see figure 8.4C). Contrast this situation to figure 8.3, which does incorporate asymmetrical relationships and in that sense presents an advance over figure 8.4. Yet, even configurations with asymmetrical relationships do not clarify why an animal with such configurations in its brain should have any "feel," or know what it is like to have its olfactory module active (cf. Nagel, 1974). Neural activity and connectivity patterns can be readily simulated on a digital computer, and when this is done, one cannot escape from concluding it would be very hard or impossible to justify that the computer is actually conscious. When we study the brain's activity patterns while different sensory organs are being stimulated, we certainly will observe modality-specific configurations. However, this approximates a truism—given our knowledge about specific sensory tuning of neurons, *of course* it is the case that resulting activity patterns will be differentiated and complex. However, why and how would this give rise to anything like consciousness (cf. Chalmers, 1995, 1996)?

Compare the brain and its high-dimensional activity patterns to a bucket filled with nitrogen, carbon, hydrogen, sulfur, oxygen, phosphorus, and some other kinds of atom. A crazy scientist then instructs a bystander by saying, "If you configure these atoms in the right way, a cat will jump out of the bucket!" One crucial step in explaining the neural basis of consciousness is to tease out *how* exactly neural activity patterns are spatiotemporally organized, at different levels of complexity, taking the biological organization of atoms into molecules, proteins, subcellular organelles, cells, tissues, and organs as a metaphor.

Regardless of how phenomenal content might be generated, we should be more precise about the requirements for systems generating consciousness. Characteristic of the experiencing of sensory inputs is that percepts are usually *unambiguous*. Also Churchland

(1995) considered this capacity for alternative interpretations an important dimension of consciousness. Although in figure 4.5 one can learn to see the textural pattern of dots and stripes without consciously recognizing the animal's identity (duck or rabbit), the normal course of affairs appears to be unambiguous interpretation. This lack of *interpretive* ambiguity is akin to *perceptual stability*, and it is interesting to explore whether these concepts as well as their neural substrates can be segregated.

In psychology, interpretation is often linked to *recognition* as a semantic memory function: at a given instant we recognize the sensory input as a duck, but then we rapidly switch to "seeing" a rabbit. We will not see a third species emerging because we have not learned about this new kind of animal before. In the psychological literature, memory-based interpretation of sensory inputs has been elaborated by explaining perception as a constructive process depending on the interaction between sensory input and stored knowledge (e.g., Bruner, 1957; Gregory, 1980; Marcel, 1983).

A general problem in establishing an interpretation is how particular sensory inputs come to be put in place as evidence: which inputs are worth considering, and which others should be ignored? A first subprocess in this selection is attention, which focuses us on the most relevant subspaces of all sensory dimensions, either in a bottom-up (attention driven by input saliency) or top-down fashion (driven by preestablished knowledge or memory; Duncan, 1984; Posner, 1994; Desimone & Duncan, 1995; Driver et al., 2001). Attentional processes should be segregated from conscious processing per se, because we can be conscious of inputs without attending to them (Braun & Julesz, 1998; Montaser-Kouhsari & Rajimehr, 2004; Koch & Tsuchiya, 2007). The two processes also appear to be neurally dissociable (Watanabe et al., 2011). Altogether, psychophysics suggests regarding attention as a function to steer and amplify conscious processing and facilitate relevant inputs making the transition from nonconscious (or preconscious) to conscious processing (DeHaene et al., 2006). Secondly, the selection of inputs as evidence for perceptual constructs relies not only on single-source evidence but also on the degree to which different sensory sources match or mismatch when converging at higher multimodal levels in cortex.

In the neural domain, recognition and categorization have been firmly associated with the temporal lobe memory system (particularly peri- and entorhinal cortex, and (para)hippocampal areas) as well as prefrontal areas (Eichenbaum et al., 2007; Freedman & Miller, 2008; Murray et al., 2007). In contrast, perception is associated with the primary and secondary sensory cortices, and especially higher associational areas such as posterior parietal cortex and posterior cingulate cortex (see chapters 3 and 6). But how solid is this classic dichotomy between "perception" versus "recognition and categorization"?

Looking at a "Rubik's Cube" (see figure 8.5, plate 9), the difference in color relates to the effect of context—formed by the neighboring bins—on the perceived color of the central patch. The related phenomenon of color constancy enables us to recognize the same patch of color when the spectral content of ambient light is changing (Land et al., 1983). An apple that looks green in bright daylight is still perceived as green upon sunset, when

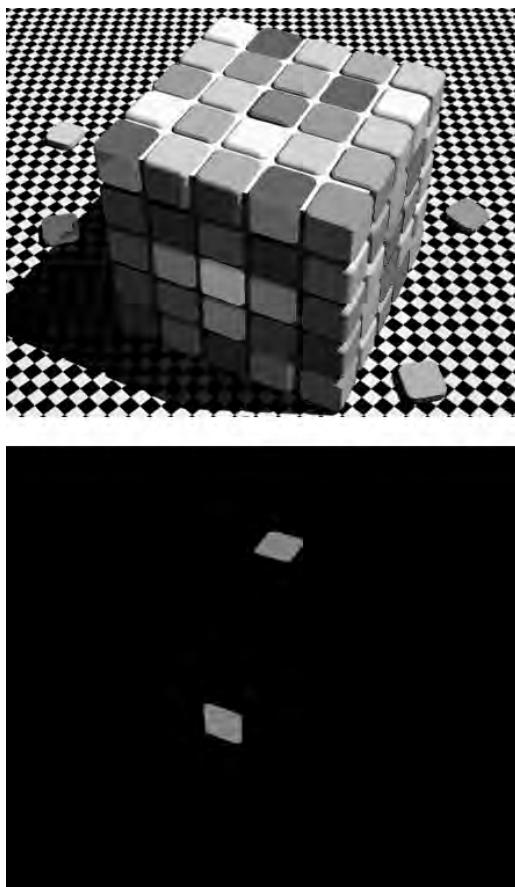


Figure 8.5 (plate 9)

Color constancy and color illusion. Focusing on the center pieces of each of the three visible sides of this “Rubik’s Cube” in the upper panel, two are brown and the third one, on the shadow side, appears orange. In the lower panel, a dark mask has been used to remove the contextual effects of the tiles surrounding the center piece on each side. Now the colors on the top and shadow side look identical. Contextual color effects result in the same patch of color being perceived as darker when the surrounding is brightly lit as compared to the shadow side. Reproduced with permission from B. Lotto.

the spectrum of the sun's light reaching the eye is shifted toward red due to the atmosphere's filtering properties.

Figure 8.5 (plate 9) amazes many of us by the way color context manipulates our perception of something we take to be solid and unmistakable. But should one call this an illusion—a percept of something which is not really there in the outside world? I would argue that both the brown and orange colors are veridical percepts and that also here our brains entertain two different *interpretations* of the central bin's color in a way that is not *fundamentally* different from a duck versus rabbit interpretation.³ The main difference lies in the extent to which the two kinds of interpretation are supported by memory.

Inconspicuously, the term “meaning” has been sneaking into our discussion along with “phenomenal content.” Let me briefly explain how the term is used here, leaving linguistics, semiotics and societal connotations aside. When we define the meaning of a word or sentence, we usually describe or circumscribe the target entity by summing up synonymous or related words. At present we are not concerned with language, but with conscious representation, including its mnemonic and categorizing components. In this domain, we first use the term “meaning” in referring to percepts or imagined sensations. For instance, when I am on the phone with my wife and mention a “cream white” color of the tiles to be used for a new bathroom, and she answers “I know which color you mean,” she refers to a retrieved memory of that color, reinstated by imagination. This first use can be captured as *phenomenal meaning* (see chapter 1; cf. Block, 1990, 2005). A second usage relates to the host of other entities and properties we have learned to associate with the object having a meaning we are referring to (see figure 5.2). When you say, “To a mouse, a cat means danger,” you are referring to an emotional contingency associated with cats and mice in general, not to a particular cat or its properties you are perceiving. *Associative meaning* (or meaning acquired by associative learning) is an appropriate term to describe this second use and is intimately related to traditional semantics or more recent variants, ascribing meaning to a term by way of its connotations to other entities or words occurring in the same sentence (e.g., Leech, 1977; for teleosemantics, see Millikan, 1984; for psychofunctionalism, Fodor, 1975). We may question whether there is a sharp border between these two different types of meaning. As for perceptual *versus* memory-driven and cognitive interpretation, I argue there is no sharp boundary between perception and cognition, as one can be viewed as a continuation of the other in different form. Contemporary neurocognitive research supports that perception is impregnated with cognition.

As a point in case, let us compare visual agnosia to akinetopsia (see chapter 3). A visually agnostic patient may have trouble recognizing and identifying visual stimulus configurations while perception of individual elements remains intact. Agnosia patients can be alternatively said not to “see” a particular input configuration as a chair (while still reporting its various line elements) or to be unable to “interpret” the input as a chair. As for the duck–rabbit, the distinction between “seeing” and “interpreting” appears fuzzy, and the choice of terms becomes somewhat arbitrary (cf. Zeki, 1993; Farah, 2004;

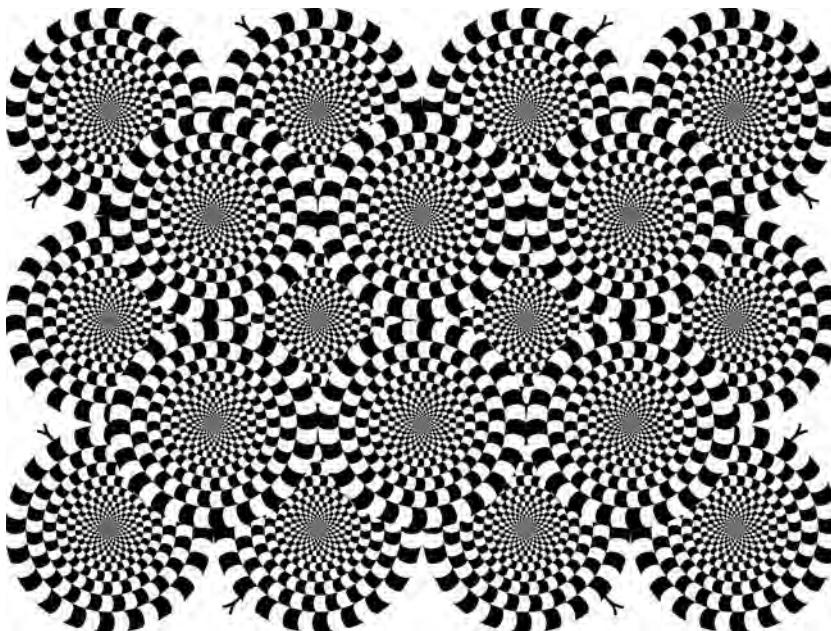


Figure 8.6 (plate 10)

The “rotating snakes” illusion of movement. Rotations are experienced most strongly in the peripheral field of vision. Large contrast differences between adjacent segments of a coil (e.g., green–white or white–blue transitions) are important for eliciting the illusion, which might be explained from small discrepancies in processing delays in the peripheral visual system. About one in four people are not susceptible to this kind of illusion, which involves a genetic basis (Fraser & Wilcox, 1979). Reproduced with permission from A. Kitaoka, Department of Psychology, Ritsumeikan University, Japan (see also Murakami et al., 2006).

Riddoch et al., 2008). However, should we not classify akinetopsia as a purely *sensory* deficit then? For most objects around us we do have a clear, all-or-none impression that they are either moving or not, and doubts only arise in cases of very slow movement. Is there any room for interpretation?

Figure 8.6 (plate 10) powerfully contradicts the idea that a perceived object is either exclusively moving or not moving; which parts of it appear to be rotating depends on your direction of gaze, with circles in your peripheral visual field having more illusory motion than central ones. Returning to akinetopsia, evidence suggests that damage to area MT/V5 lies at the basis of this disorder (in humans: lateral occipital cortex; Schenk & Zihl, 1997; Zeki, 2001). When area MT/V5 is selectively lesioned in monkeys (Newsome & Paré, 1988), motion vision is heavily impaired, but most likely a subset of V1 cells will still be sensitive to motion. A subset of cells in the “magnocellular” pathway from the LGN, via layer 4B of areas V1, V2, and V3, to MT fires selectively to motion directions of visual stimuli (DeYoe & Van Essen, 1988). MT neurons have much larger receptive fields than V1 neurons, as they globally integrate motion cues across a wider visual field and also

include disparity and other trajectory information. Therefore, the situation of akinetopsia caused by MT damage resembles that of visual agnosia: low-level sensory elements of visual motion will still be processed, like the line elements of a chair, but a higher-level interpretation of the entire object as “moving” is lacking.

Vice versa, it is not farfetched to view a classically “cognitive” process such as recognition as being part of, or contributing to, conscious perception. The study of object recognition is usually operationalized by asking subjects whether they have seen an object or person before or not, amounting to a binary novelty-versus-familiarity decision. Watching a person walking past you in the street may evoke a strong feeling that you have seen that person before although you cannot name him or describe where he is from. The sense of familiarity can be so strong that you can almost smell it. Recalling that subjects with medial temporal lobe damage are still conscious, we also acknowledged that this consciousness, devoid of declarative memory, is impoverished (see chapter 3). There is no a priori necessity to classify the sense of familiarity purely as a cognitive function, as it may act as an additional “sense,” accounting for background, quasi-emotional feelings we experience when meeting a person. In other words, familiarity can be viewed as a higher-order, parasensory modality. In line with this, the perirhinal cortex occupies a higher-order position well-integrated into the sensory processing hierarchy and has been implicated in coding novelty–familiarity judgments (see figure 6.3, plate 3; Murray et al., 2007; Eichenbaum et al., 2007; O’Neil et al., 2012; chapter 3).

The bottom line, then, is to incorporate interpretation into a set of fundamental requirements for consciousness. Interpretive processes proceed at a “higher,” more cognitive level, as illustrated by visual agnosia, or occur at lower levels, closer to the entry site of sensory input into thalamocortical systems. In contrast to the “modular mind” view of psychofunctionalism (Fodor, 1975; Pylyshyn, 1999), this implies we are no longer maintaining a sharp boundary between “perception” and “cognition”: both are fundamentally a form of *gnosis*. Interpretation assumes grades of complexity, depending on the position of the processing node in the sensory–cognitive hierarchy.

This is not to say that sensory input can be dispensed with as soon as a system achieves interpretation. The brain continuously samples sensory inputs, performing reality checks on ongoing changes. When you approach your car parked on a street, you have a rough idea of what the side of the car that you cannot see will look like. In theory, your interpretive system could retrieve these features from memory and “fill them in” to construct a complete, three-dimensional percept, yet you do not see the car’s other side (cf. Jackendoff, 1987). When we think of the brain as a world-modeling device, it is crucial to maintain a distinction between actual sensory input and what the world is predicted to be like. Consciousness is chiefly concerned with the actual situation. The lack of three-dimensional “filling in” is functional, because a direct visual input is lacking and there might be unexpected objects at the other side, such as a hole in the street or scratch on the door.

8.5 Some Remarks on Self-Awareness

Consciousness is sometimes taken to include the “self” as one of its inalienable dimensions. I will argue here in favor of maintaining a clear distinction between consciousness and self-awareness. The main rationale is that consciousness does not depend on an active notion of “self” as there are many instances of conscious experience where this notion is lacking or where a self-identity is not unambiguous or clear. That the “self” is not a separate entity distinct from body sensations is a view standing in a long tradition, beginning with David Hume (1711–1776) and contemporaries. Hume (1739) argued that our experiences of our “self”—even when we label these as “introspective”—do not stand apart from our body and its sensory–cognitive capacities, including perception, memory, and thought. More recently, Damasio (2000) proposed that consciousness critically depends on the genesis of a nonverbal account of how an organism’s own state is affected by the processing of external object (i.e., nonself) information. He builds this argument on phylogenetically ancient brain stem mechanisms that regulate the balance of needs and consumptions in our internal milieu homeostatically and refers to this foundation as a (preconscious) “protoself.” In chapter 6 I also placed these mechanisms outside the domain of conscious processing, let alone self-awareness. Let us start reexamining the roots of self-awareness by considering different notions of “self.”

In daily life, we entertain different yet related notions of the self: we realize that our “I” is socially distinct from other agents around us, that we each have a body distinct from an external environment, and that we refer particular sensations to this body. We talk about our “self” as an entity that makes us behave spontaneously and takes decisions that are not obviously and immediately dictated by the environment (Metzinger, 2000, 2008; Churchland, 2002). When restricting our view to consciousness per se, however, the notions of our “social self” and “self as agency” are less relevant. Perception is not logically and practically dependent on the representation of social relationships or on the presence of conspecifics in our environment. Nor is consciousness necessarily dependent on volition, planning, or decision making, as illustrated by patients with severely impaired motor and prefrontal systems. This is not to say that social relationships or volition would be unimportant for self-awareness; the only claim is that awareness—regardless of self–nonself distinctions—is not fundamentally dependent on these factors.

While daydreaming, one can stare in the distance, perceiving objects such as houses, bicycles, and cars. When doing this, I only realize slightly later that it is *me* who is seeing all that, situated at the place where I happen to be. Self-awareness appears to follow general perceptual awareness although one may object that the “self” had been sitting at the fringe of consciousness from the very start. If we adopt an ontogenetic perspective and take various developmental tests in children about 1,5 to 2 years of age as indicators of their self-consciousness (such as looking in a mirror and touching a mark placed on the child’s own head), they indicate experiences devoid of a self–nonself distinction earlier than when

talking about, or referring to, themselves as entities distinct from their environment (e.g., Lewis, 2003).⁴

Also neurological patients provide striking examples of how consciousness can be dissociated from the self. Recapitulating out-of-body experiences, it seems hardly possible to argue against the idea of a unitary self more clearly than by citing the reports of patients seeing themselves lying down below in a bed while they are undeniably conscious and have a single, unified view. However, delusions of the self are not limited to these cases. Patients with right-hemisphere brain damage, usually involving parietal and insular areas, can have a persistent sensation that one of their limbs does not belong to them and is not an integral part of their own body—a syndrome called “somatoparaphrenia” (Vallar & Ronchi, 2009). This distorted sense of body “ownership” is expressed, for instance, by a patient’s lying on a bed and trying to push his right leg out of the way, as it is experienced as an external, alien object.

Distortions of self-perception have been reported by people taking psychoactive drugs. Ketamine is known as a dissociative anesthetic: subjects in a “K-hole” describe some of their psychedelic sensations as becoming detached from their own body. They report out-of-body experiences or feel that they “melt into the surroundings,” accompanied by a strongly distorted sense of space and time (Curran & Monaghan, 2001).⁵ Two ketamine users gave the following accounts (Muetzelfeldt et al., 2008):

“Everything was surreal, like in a daydream where nothing was very real. I couldn’t feel my body; it was different from anything I had tried before.” [...] (p. 221).

“I really liked the trip [...], it was very intense. Had an out of body experience and it’s like there was a whole new world in your head and a lovely wave of relaxation washed over you.” [...] (p. 221).

Several chronic psychiatric disorders involve distorted self-perception and self-control, including schizophrenia and multiple personality disorder (also dubbed “dissociative identity disorder”; Putnam, 1993; Reinders et al., 2003). Because these disorders have a highly complex etiology, it is difficult to pinpoint how specific self-misrepresentations come about. Nonetheless, interesting parallels can be drawn to other disorders malforming the sense of self. A hallmark of schizophrenia is vigorous hallucination, often expressed as auditory sensations that seem not to be generated by the patient but are misidentified as originating from an external source (“I constantly hear this voice in my head that tells me what to do”; Woodward et al., 2007; Sommer et al., 2010). fMRI, MEG, and EEG studies in patients have shown that the left primary and secondary auditory cortices are hyperactive during auditory hallucinations (Dierks et al., 1999; Ford et al., 2009; Van Luterveld et al., 2011). Concurrently, the medial prefrontal cortex shows deficient activity, which is likely to result in dysfunctional cognitive control and reality monitoring (Vinogradov et al., 2008). The combination of these two changes may result in a situation where the patient experiences conscious auditory percepts but lacks the capacity to steer and refer the flow of

these auditory percepts, in contrast to auditory imagery in healthy people. A sense of ownership will be lacking, and the percepts are incorrectly attributed to an external source—be it a chip implanted by the CIA or the voice of a zealous leader. Under this working hypothesis, dysfunctional self-representation is related less to sensory representation and more to the “self” as an active, volitional agent. However, regardless of whether the hypothesis is correct, schizophrenic hallucinations reinforce the point that conscious experience continues under conditions of a distorted self.

When considering self-awareness in relation to the “protoself” circumscribed by Damasio as the basic homeostatic regulatory systems that map and safeguard the integrity and health of the organism, I argue we should regard these systems, although vital for body survival, as being separate from consciousness or self-awareness.⁶ “Consciousness” means, in the first place, *perceptual* consciousness and has to meet more requirements than those needed to maintain cell and body function. Self-awareness comes into play at a more advanced stage than consciousness, as a form of metacognition. The realization of an “I” follows the emergence of conscious representations in time, ontogenetically but often also phenomenologically. As applies to consciousness in general, self-awareness depends on the seamless cooperation between different sensory mapping systems. Some of these systems are structured as a body map, but equally important are the frameworks used for exteroceptive senses, vision and audition, because the comparison between body and exteroceptive maps is prime to distinguishing the “self” and the “nonself,” the outer world. When all these mapping systems with distinct frameworks smoothly collaborate, we have the experience of a unified perspective and unified self—as there is no subsystem that would signal discrepancies. It is only when our mapping systems produce conflicting pieces of evidence that we notice our self is not so monolithic as our language suggests it to be.

8.6 Covert Dualism: Pitfalls in Describing Conscious Representation

The notion of “self” is so deeply interwoven with the fabric of our thoughts that it is hard to articulate concepts of conscious representation completely devoid of it. Yet, this is exactly what we should be doing. If we explain a conscious representation in terms of a “scene,” for instance, we implicitly assume there is *someone* looking at that scene, an “I” who has a view on that scene. If our thinking perseveres on that stance, we are doomed to dualism, either overtly or covertly: we maintain a separation between our brain, or representational system that offers the materials, and some mental entity that does the job of viewing and interpreting. Theories of mind–brain have proposed concepts such as a “sketch” (e.g., the 2.5 D sketch of Marr, 1982) or “theatre of mind” (Baars et al., 2003), but the misleading aspect of these metaphors—which are invoked almost inevitably—is that they presume some kind of spectator or homunculus viewing the materials on display. The same risk is borne out by Crick and Koch’s (2003) proposal that “the front of the brain is ‘looking at’ the sensory systems, most of which are at the back of the brain” (p. 120). If Dennett

(1991, 2005) were not famous for combatting hidden Cartesian implications, one could even argue that his multiple drafts model hinges on a covert persona reading and editing the drafts—although he is careful enough to point out that this would be a misconception.

One may object that also an immersive, multimodal–situational representation entails the concept of *somebody*'s being in that situation (although “looking” or “viewing” is less strongly implied here—and certainly not by necessity). When describing experiences, personal connotations are unavoidable because of the way our language has developed and is invariably used with reference to individuals. My goal here is to clear our concepts of consciousness as much as possible of these connotations. Whenever you will see the use of “scene” or “situation” continued to denote a conscious representation in the current discourse, it is meant to be a self-sustained event, without further requirements for observers or interpreters outside the system for representation. If we were contemplating a multi-area brain system and somebody forced us to tell which of the areas acts to *observe* the perceived elements, the best answer we may offer is this: all areas that *generate* these elements. The observer is the producer. In the next chapter I will argue that the mere presence of essential nodes is not sufficient for perception. Its generation depends on many areas working in concert, and a conscious representation should not be attributed to any single area in particular.

In summary, brain systems must meet several hard requirements to qualify as conscious, while other requirements are “soft” in that they are important for sustaining normal, daily-life awareness but not strictly necessary for having a basic form of conscious experience. The hard requirements include the ability to interpret (or reconstruct) sensory inputs as having particular qualities or content, within a rich repertoire of (sub)modalities. Secondly, this process of attributing sensory “feel” or meaning to inputs can occur in a dynamic or stable state, depending on the constancy of variables governing the sensory flux. Projection of interpreted sensory inputs into an external, perspectival space (vision) or body map (somatosensation) is seen as a relatively basic process, but patient studies indicate that the core of consciousness does not strictly depend on this ability, as applies to normal requirements on the grouping of similar features and binding of different submodalities into objects. Also the “unity” of consciousness and self-awareness are not classified as an essential feature but rather as a constantly maintained “illusion” of the healthy brain empowered by proper multimodal and motor alignment.

9

Neural Mechanisms for Conscious Representations

I like cabernet sauvignon. More specifically, I like the taste of this wine. But what I like is not the mere existence of cabernet sauvignon or its taste; what I like is *experiencing* its taste. But is even this a satisfactory characterization of what I like? Let's suppose for a moment that the taste of wine is some chemical property of it; that what one is detecting when one experiences that distinctive cabernet sauvignon taste is some complex combination of esters, acids, oils, etc., although of course one does not experience it *as* having such a chemical composition. I suppose that I might learn to detect this same property visually, with the aid of the apparatus in a chemistry laboratory, and so have an experience whose intentional content is the chemical property that is in fact the taste of cabernet sauvignon. But when I say that I like the taste of cabernet sauvignon, or like experiencing this taste, I don't mean that I would like having such a visual experience.

—S. Shoemaker (1990)

9.1 Neural Mechanisms for Coordination between Mapping Systems

In this chapter we will enter the difficult subject matter of the neural mechanisms that may fulfill the key requirements for conscious representations, which calls for more theory building than in previous chapters. Following a discussion on functional convergence of spatial mapping systems in the brain, we will look at low-level integration processes such as feature binding. This will lead us to ask whether sensory feature detectors of varying complexity meet integrative requirements at a single-cell level and whether, and why, *connectivity* between detectors is essential. Addressing higher levels of integration, we will approach questions on interpretation by returning to the problem the labeled-lines hypothesis left us to resolve: *modality identification*. Whereas modality identification implies a question of sensory interpretation, the problem of how we apply acquired knowledge to sensory inputs poses the related question of *cognitive interpretation*. Here we enter the minefield of how phenomenal meaning arises and how we come to interpret sensory content unambiguously. Altogether, these deliberations converge on the current theory's central tenet that a multitude of sensory–cognitive architectures must actively interact to form and maintain conscious representations. As yet, this yields a static diagram of interactions,

and so we need to address how the system can both accommodate salient environmental changes and maintain a stable representation when such changes are subtle.

Our sensorium contains a wealth of mapping systems, often set in different coordinate frames. The availability of maps per se is not sufficient for sustaining consciousness because, for instance, they are also found in areas associated with nonconscious processing such as the cerebellum. Under normal circumstances we visually perceive the world from a first-person perspective, but having a spatial perspective or spatial mapping is not exactly a necessary condition for consciousness per se (see chapter 8). The main problem concerning perspective is how to make different multimodal mapping systems collaborate smoothly.

When scrutinizing neural mechanisms potentially accounting for the impression of perspectival unity, it appears unlikely that different low-level maps in multiple sensory modalities would converge onto a single, higher-order map. Apart from the issue of whether a single, overarching map exists, multimodal convergence subserving spatial referencing likely takes place in, at least, a set of parietal and frontal (particularly premotor) areas (Goodale & Milner, 1992; Galati et al., 2000; Committeri et al., 2004; see chapter 3 and Karnath et al., 2001, for temporal areas). Specifically, convergence of visual, auditory, and somatosensory information in the dorsal, “spatial” stream is thought to occur in the inferior parietal lobule, which includes parietal area 7A and the LIP (see chapter 6; Xing & Andersen, 2000), and in ventral premotor cortex (Graziano, 1999; Lloyd et al., 2003). Multimodal information also converges onto the dorsolateral prefrontal cortex, but lesions in this area cause primarily working memory and related executive deficits, not loss of consciousness.

It is still unclear whether spatial alignment of multimodal inputs to a common reference frame—or conversion into a common perspectival “format”—actually takes place in the brain. Would this alignment be strictly needed to experience a coherent representation of the world? Spike recording studies in parietal areas 7A and LIP, supplemented with network simulations, suggest that representing all information in the same reference frame is unlikely to be necessary per se to achieve coherent spatial representations and ensuing motor patterns (e.g., Andersen et al., 1985; Zipser & Andersen, 1988; Snyder et al., 1998). Even within a single area of the parietal cortex, different cells can obey distinct frames of reference. In macaque monkeys performing a task requiring gaze shifts to different stationary targets, Avillac et al. (2005) found multimodal cells in the ventral intraparietal area (VIP) responding in different frames of reference, varying from eye- to head-centered. The coexistence of multiple frameworks does not hamper execution of the task and was argued to allow input from one modality (e.g., vision) to predict stimulation in another modality (e.g., somatosensation). That these visual inputs can be aligned or misaligned relative to tactile inputs is apparently not in conflict with a normal functioning of the system. In a related study (Pouget et al., 2002), a model was presented showing how visual, eye-centered information can be combined with eye-position information to yield a coordinate transformation into head-centered information, underscoring that it is relatively

straightforward to combine and align different reference frames in a limited neuronal network (see figure 9.1).

In line with these experimental studies by Duhamel's group, Deneve and Pouget (2004) argued on computational grounds that the conversion of differently aligned, multimodal inputs to a common (motor) output can be solved by populations of neurons having different degrees of (mis)alignment (see figure 9.1). Although these studies, altogether, do not explain why having one dominant perspective is characteristic for visual experience,¹ they reinforce the thesis that no common egocentric mapping exists across sensory modalities; a first-person perspective is subjectively experienced but arises nonetheless from cooperating yet differently referenced mapping systems.

Nevertheless, some differences between mapping systems are more important than others, particularly when distinguishing the external world from our "self," our own body. In the somatosensory domain, touch, heat, pain, and other sensations are referred to the homunculoid map in Brodmann's areas 1, 2, 3a, and 3b (see figure 2.11C). We do not need to compute our body orientation relative to external objects to render this projection correctly because this external information is irrelevant for precisely localizing a sensation on one's body. This situation contrasts sharply to the visual and auditory system, where our capacity to localize a sound or light in external space must take into account how our body, head, and eyes are oriented in space in order to produce an allocentric (world-referenced) representation. If we are asked to point to a sound source somewhere in a room, we can effortlessly point to the same spot when our head has rotated to face the sound source directly (see figure 9.2), even though the critical sensory input for source localization differs (Altmann et al., 2012). When we glimpse an object lurking in our visual periphery and shift our eyes to foveate it, light reflected from the object hits different locations on our retina, but our brain localizes the object at the same position in space as it corrects for eye and head movement (see figure 9.1; Snyder et al., 1998). The same can be said of gaze fixation of external objects as we are walking. These evolutionarily ancient mechanisms make us distinguish localized sensations on our own body from external objects, whose localization requires correction for body orientation.

Taking this a step further, we may postulate that the projectional nature of conscious representation—the referral of sensations to "things in the outside world"—arises from this subtle, "undercover" collaboration between visual, auditory, vestibular, somatosensory, and motor systems. Importantly, motor systems inform sensory systems about body movements via efference copy and proprioceptive feedback (Jones et al., 1978; Stepniewska et al., 1993; Nelson, 1996; Whishaw & Brooks, 1999). Our fundamental "sense of space" and experience of a world external to ourselves does not need to be shaken by this, but it is worthwhile realizing it rests on a sensorimotor construct, a conspiracy between sensory and motor mapping systems that afflicts us with the "illusion" of perceiving an "objective," nonself space. We watch an apple lying on a table and interpret the apple and table as being external to ourselves because—even when we walk around the table, turn our head, and

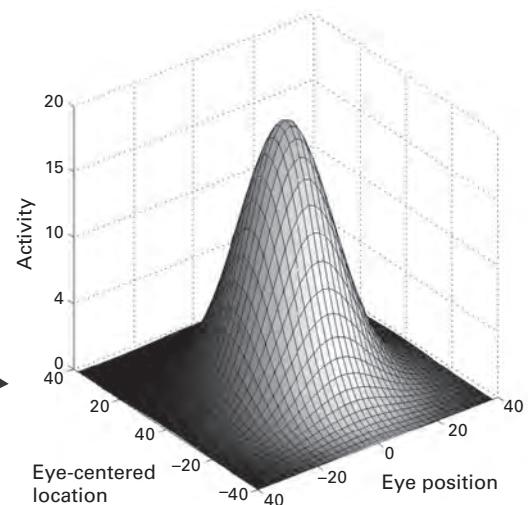
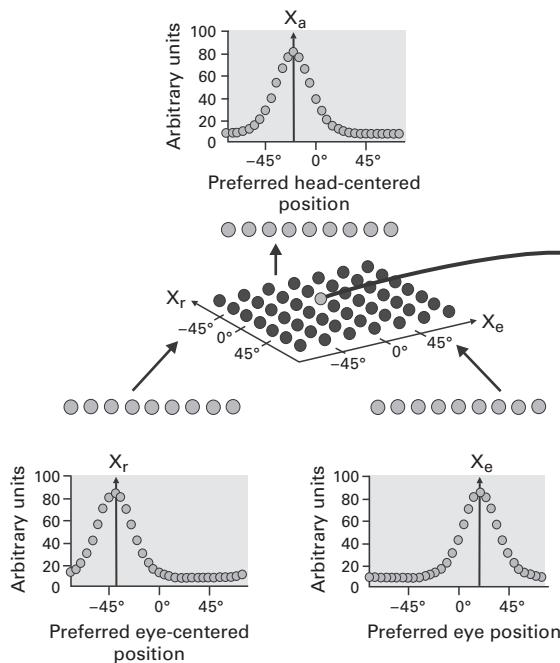
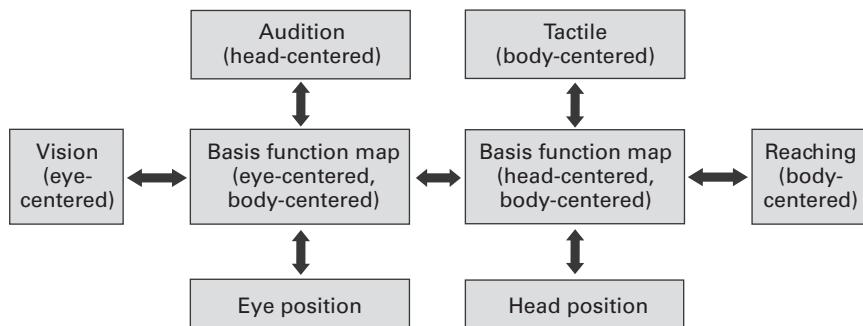
A**B**

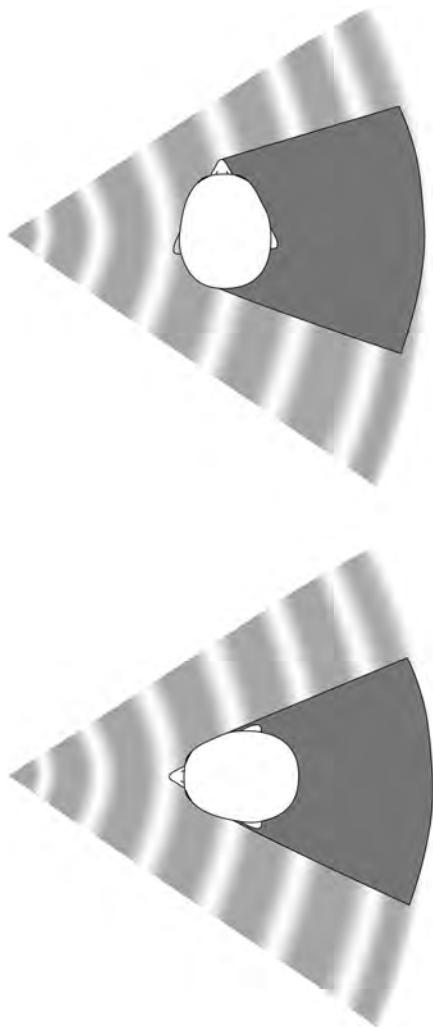
Figure 9.1

Neural network simulations suggest that representing all information in the same reference frame is not necessary to achieve coherent spatial representations and corresponding sensorimotor behavior. (A) A network that coordinates transformations of information coded in different reference frames by neurons coding basis functions. This network tackles multisensory spatial integration by converting two different inputs (coded by the two input layers at the bottom, using eye-centered information and eye-position information, respectively), set in different reference frames, into the activity pattern of an intermediate layer, combining the two inputs. Next, the network outputs information in yet a different reference framework (head-centered, top). Each input layer consists of a retinotopic map of 32 units (only nine units shown for each layer). The term “eye-centered” means that receptive fields are anchored to the eyes. When the eyes move, the receptive fields shift by the same amount. We suppose that a visual object is presented at location x_r , defined in eye-centered coordinates. Neurons in the left input layer that are tuned to this position will fire strongly, according to the Gaussian curve at the bottom. Similarly, the right input layer will generate a population code for the eye position x_e , given the object location. These two population inputs converge on the cells in the intermediate layer, which are sensitive to both eye-centered location x_r and eye position x_e . This joint sensitivity is expressed for one neuron in its basis function (right panel, two-dimensional bell shape), a concept referring to vector metrics. Multiple vectors are said to form a basis when any other vector can be constructed as a linear combination of these vectors. Each intermediate unit computes the product of the activity from a pair of eye-centered and eye-position cells. The peak in the graph on the right shows the preferred eye-centered location and eye-position combination for an example intermediate unit. When the activity of the intermediate units is linearly combined, the firing activity of the output units, set in head-centered coordinates, is computed. As a result, the visual object location is now coded in a craniotopic reference frame: when the head rotates, the receptive fields shift by the same amount, whereas eye movement is corrected for due to the processing of eye position. This model illustrates one of several contemporary approaches to sensorimotor transformations (see also, e.g., Zipser & Andersen, 1988; Mazzoni et al., 1991; Pennartz, 1997). (B) Multimodal basis function network for converting sensory information set in different reference frames to other coordinate systems, used both in the sensory and motor domain. Note that this network has connections in all directions, allowing predictions to be made from one modality to another, including the motor domain, as well as estimations of the sensory consequences of motor actions. In this scheme, some basis function maps encode auditory and tactile targets in visual coordinates. Adapted from Pouget et al. (2002).

move our eyes—our brains compute these objects as remaining at the same location. I use the term “illusion” here not to imply that our experience of the apple would be false but to emphasize the constructive nature of spatial experience.

That spatial mapping systems cooperate seemingly without effort in the healthy brain by no means implies these operations can be carried out on the fly. In the model of Deneve et al. (2001), for instance, the connection weights between retinotopic visual and eye-position inputs and centrally placed neurons are adjusted by training to avoid a malfunctioning conversion of multimodal inputs to motor output. Such operations begin during early development, as infants learn elementary visuospatial tasks such as eye–hand coordination (Von Hofsten, 1982). Feeling a scratch on your skin and seeing that same spot on your leg requires a cross-modal identification enabled by prolonged practice. The spot on your body map associated with itch needs to be matched to a visual localization, enabling the brain to integrate these two sensations.

What may help the brain forward during such early learning is our motor system, acting as the “Great Calibrator.” Consider the awesome toddler learning to walk. She steps forward toward a touchy-feely apparatus that will start playing the “Old McDonald Had a Farm” tune when touched. This single step results in a bewildering multitude of sensory

**Figure 9.2**

Localization of a sound source given variable head orientation. Top: A speaker, positioned to the left of an individual who faces a 90° direction away, generates sound waves that first reach the left ear. The brain infers the position of the speaker using the time delay caused by the sound wave's longer travelling time to the far, right ear (interaural delay), as well as the decrease in intensity of the sound waves reaching the right ear (shadow). Bottom: Same as top panel, but now with the person facing the speaker. Despite the very different delay and level differences, the brain infers the same sound source location given knowledge of the person's head rotation. The picture illustrates especially that sounds are easily perceived within a world-centered (allocentric) framework, unlike the “egocentric” first-person perspective often emphasized in vision.

changes: the sound becomes louder, the machine looks bigger and is suddenly within reaching distance of her hand, and the step movement itself is sensed by proprioceptive feedback. All of these sensory changes relate to a common event—the forward step—which can thus be used to “calibrate” them relative to each other and reduce potential mismatches between multimodal inputs.

In short, single sensory neural maps have a spatial layout, but apart from this they do not have, or embody, any “intrinsic” knowledge of space—the spatial aspects of our experiences are not inherent to having a map set in some coordinate system. It is only through an elaborate process of development and early-life experience that different sensory and motor systems come to cooperate and enable us to distinguish “self” from “nonself,” internal from external. Although in adulthood space nearly qualifies as a Kantian a priori attribute of experience (see chapter 8), the current theory predicates this condition on developmental processes, once again making it less absolute.

9.2 Neural Mechanisms for Integration: Feature Binding

Given all of these distributed maps, how is it that different modal features belonging to the same object are grouped together and do not come to be assigned to different objects? We already encountered the binding problem (see chapters 3 and 7), and it presents the first low-level stage of integration we will study. When we go further back in time than the 1970s, feature binding was not recognized as a genuine problem. Departing from a plethora of low-level *feature detectors*—sensory neurons specifically responding to a narrow range of some feature parameter such as color, shape, or taste—Barlow (1972) reasoned that information from multiple low-level visual detectors can be combined to excite a small group of high-level feature detectors. This process of combination by convergence can be pursued to yield the ultimately specific neuron—the cell that would only be activated by the sight of one of your grandmothers (the “grandmother cell,” a term coined by Jerry Lettvin; Gross, 2002; see figure 9.3).

Barlow’s combinatorial scheme led him to propose that the brain represents stimuli using a minimal number of specifically tuned, active neurons (“cardinal cells”). This way, he made sure that features belonging to the same object are represented together: take care that information about multiple features all converges to object-specific neurons, and there you have a neural substrate for feature integration.

However, one problem with this combinatorial scheme is that perception of different object features greatly depends on anatomically segregated brain areas (see chapter 3 and 6). Moreover, even given the huge number of neurons available in the visual system, there is no sufficient amount to represent all combinations of features, either as previously experienced or considered as the complete feature space of all possible groupings. When you think about the faces you have seen in the past, the estimated number you have stored may not seem prohibitive at all relative to the number of neurons in your visual system.

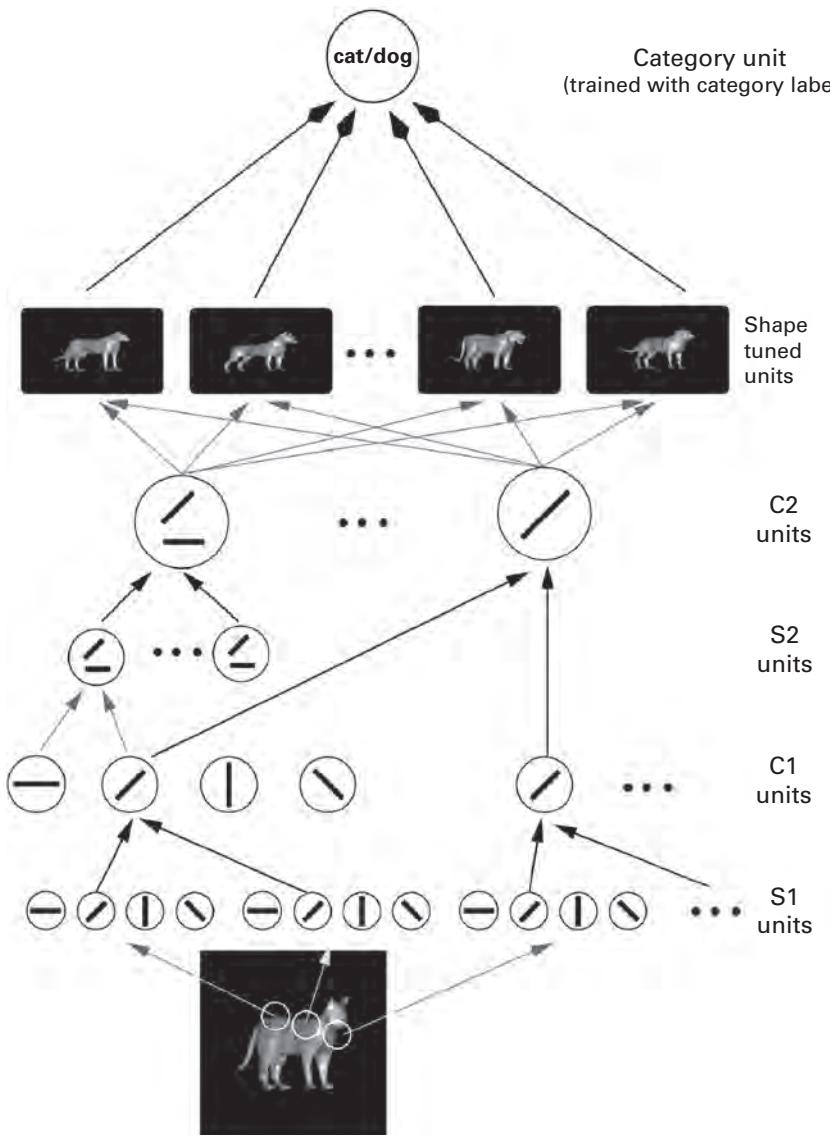


Figure 9.3

Neural architecture for object recognition, inspired by Hubel and Wiesel's (1962) work and the grandmother-cell hypothesis. The scheme is based on Riesenhuber and Poggio (1999) and consists of a hierarchy of layers, starting with simple feature detectors at the bottom (S1 units) that respond to local line orientations constituting a small part of the input (cat picture). The hierarchy progresses to more complex detectors toward the top (e.g., C2). By convergence of afferent input to neurons in higher layers, simple feature responses are combined to construct tuning specificity for complex properties. The "S" and "C"-type layers perform different computational operations that ensure pattern specificity (gray connection arrows) and invariance to translation and scaling (black arrows). This combinatorial scheme results in shaped-tuned units that are both tuned to specific shapes but also keep on responding when the preferred shape is translated or rescaled in the presentation plane. Such units resemble cells found in inferotemporal cortex. Operations at still higher levels have been proposed to render the representation more view invariant (i.e., object tuned; not shown). The highest level serves to make a category decision about the nature of the object shown (cat or dog), also enabling behavioral response systems. From Freedman et al. (2003), with permission from the *Journal of Neuroscience*.

However, if we begin to imagine novel combinations such as a pink face with an elephant's nose (the Hindu god Ganesha), we realize this process may go on forever: the brain will not have all required cardinal cells ready for you, even before you started fantasizing. This issue is also known as the problem of combinatorial explosion (cf. Von der Malsburg, 1981, 1999; Churchland & Sejnowski, 1992; Gray, 1999). A further problem for purely combinatorial solutions is that neurological patients have been described having difficulty in feature binding when multiple objects are present in the situation, whereas they have no trouble perceiving single, complex objects (Treisman, 1998; Ward et al., 2002).

Adding to this, a grandmother or cardinal-cell doctrine would render the visual system vulnerable to perceptual loss, as small groups of cells may be easily damaged. In comparison, parallel-distributed networks are highly flexible, robust, and tolerant to damage (see chapter 4). However, does the very occurrence of inferior temporal cells responding to sophisticated combinations of visual features (Tanaka, 2003; Rolls, 2012; figure 3.7) not speak in favor of cardinal cells? Despite their parallel-distributed nature, model networks contain both broadly tuned, less specific cells and cells sensitive to complex feature combinations, so the very finding of complex feature detectors does not particularly favor the cardinal-cell view (cf. McClelland et al., 1986; Young & Yamane, 1992).

A scheme that attempts to explain perception from stand-alone feature detectors also faces the fundamental problem of modality identification and phenomenal meaning: detectors may fire as they like when activated by specific input, but this firing does, as yet, not *mean* anything to the system these detectors are part of. This problem is different from the binding problem, addressing which features belong together or should be separated, regardless of their meaning or content. Nonetheless, the problem of meaning adds further ground to argue against the idea that detached (unconnected) feature detectors (see figure 4.10) form a sufficient basis for conscious representation: we need to assume that feature detectors not only receive low-level sensory information in a bottom-up fashion but also are connected to, and communicate with, each other, directly or indirectly. By consequence, many of the issues having to do with conscious representation are problems of connectional organization.

Let us first look at network mechanisms that may account for feature binding. Chapter 7 reviewed the initial evidence for solving the binding problem by neural synchrony and the criticisms raised against it. Thus, it is time to contemplate alternative mechanisms for binding. A first alternative solution is to use spatially focused attention. Even before the binding-by-synchrony ideas had been tested, Anne Treisman and colleagues at Princeton University (Treisman & Gelade, 1980; Treisman, 1998, 1999) suggested that object features may be spatially "tagged" via attentional processes. When you look at a display of visual items with two relevant properties (e.g., color and line orientation) and you perform the task of searching for a specific conjunction of these features ("Find the items made up of red, horizontal lines"), Treisman posited that this feature-binding task can be solved by allocating attention to a specific region in the visual field at any given time (see

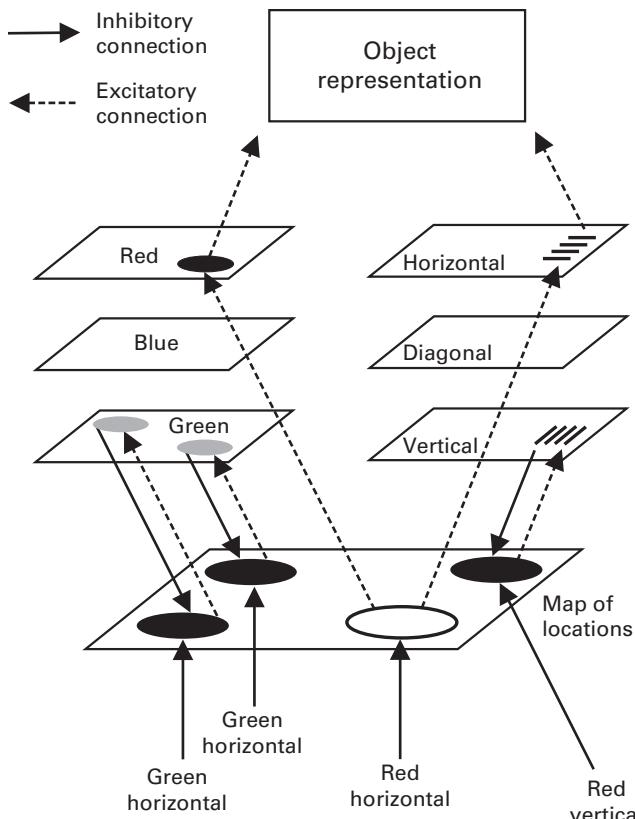
figure 9.4). This spatially focused attention at a relatively low level of the visual system (e.g., V1, V2) leads to the activation of neurons coding the presence of specific features at that location, and subsequently such “flags” for feature presence can be integrated at a higher level of object representation. Neurophysiological and neuropsychological support for spatially focused attention as a mechanism for feature binding and perceptual grouping has been raised by several labs (see Friedman-Hill et al., 1995, and Treisman, 1998, for a patient with Bálint’s syndrome, suffering from a combination of visuomotor deficits and simultanagnosia, that is, the inability to see multiple objects within a scene at the same time; see also Reynolds & Desimone, 1999; Wolfe & Cave, 1999; Croner & Albright, 1999; Shafritz et al., 2002; Roelfsema, 2006, for other relevant studies).

Yet a different way of binding features in a “top-down” fashion is memory-guided feature grouping (Peterson & Enns, 2005). Picture a brown toad sitting on a background of dead leaves in near perfect camouflage: your memory of previously seen toads will aid in completing the contours of the exemplar you are trying to figure out. Recognition of the toad is aided by the reactivation of memory traces induced by presentation of loose visual features, followed by further retrieval of object information.

There is more to feature binding than attention- or memory-based mechanisms. Psychophysical experiments suggest that feature grouping and scene segmentation can also proceed in a “bottom-up” fashion, without attentional or mnemonic guidance. These processes may be directly driven by early analysis of sensory input structure, occurring in parallel across a visual or auditory scene and even before focal attention or memory guidance can take effect. This bottom-up processing is also referred to as pre-attentive Gestalt grouping, as it performs grouping or binding based on early analysis of Gestalt features such as common motion, colinearity, and similarity (Julesz, 1981; Duncan, 1984; Nothdurft, 1992; Desimone & Duncan, 1995; Sugita, 1999; Driver et al., 2001; Roelfsema, 2006; De Sanctis et al., 2008; figure 4.6). If a bird suddenly flies right past your head, its commonly moving features are so salient and intrusive that no top-down guidance will be required.

One Gestalt feature that is tremendously useful for grouping is sometimes overlooked because it is so straightforward: the fact that features belonging to the same object usually share a common *place* within the visual field—for instance, a monkey sitting behind the foliage of a tree. This relates, of course, to the spatial component of Treisman’s model (see figure 9.4), but the point can be generalized by removing attentional requirements. Trivial as binding by location might seem, it does require that spatial maps coding different attributes (e.g., color, shape) hook up to one another to establish whether feature locations are *congruent*. Sometimes a spatial location cue is not sufficient for distinguishing objects, such as when they overlap in space. Here, common motion and texture of different body parts of the toad, sitting on dead leaves, will enable low-level grouping mechanisms to isolate the animal from its context.

Apart from the problem of which bottom-up sensory cues or top-down streams guide binding, we face the question of how binding is neurally expressed. If firing synchrony

**Figure 9.4**

Conceptual model for binding of visual object features using selective spatial attention. The model consists of a lower neural layer which codes a detailed map of locations in the visual field. The three intermediate layers represent high-order visual areas that code separate object features by way of neurons with large receptive fields, lacking spatial acuity. The three left-hand intermediate layers code for “red,” “blue,” and “green,” whereas the three right-hand layers code the orientation of the visual object presented. The upper level represents objects as entities having their features bound together (in this case, “red” is integrated with “horizontal line orientations”). The model roughly works as follows. The visual system is assumed to code one object at a time, and this object is selected by its location in the visual field (bottom layer, with neurons having small receptive fields). The bottom layer codes the locations of activated visual field regions but does not code the features that define the object present at that location. Instead, these features are coded by the intermediate layers, which can signal whether the feature is present anywhere in the visual field (“flagging”). To integrate the “what” and “where” aspects, the model posits that a window of attention moves across the location map. The attended spot then activates corresponding locations on the feature maps to signal features pertaining to the attended location. This feature information is propagated to the currently active object representation in the top layer. In the situation illustrated, an object with red, horizontal features is attended via the bottom layer (open ellipse) and the attentional window limits the upward excitation and activation patterns in the intermediate layers to the properties “red” (black ellipse) and “horizontal,” the codes of which are then combined at the top level. In a search task for feature conjunctions (such as “Find red and horizontal objects, but not red and vertical, or green and horizontal objects”), the model posits additional inhibitory connections from the feature maps downward, which serve to suppress activity at those map locations where nontarget conjunctions are presented (e.g., “green horizontal,” black ellipses on bottom layer). Adapted from Treisman (1998).

is not the long-sought mechanism to solve the problem, would ordinary changes in firing rate do? Recent experiments studying multiunit activity in primary visual cortex of monkeys performing contour grouping during a curve-tracing task indeed suggest that at least attention-based grouping can be mediated by firing-rate changes (Roelfsema, 2006). However, this does not imply that firing-rate coding would be the sole cure. If we pick out an arbitrary V1 neuron and observe that its firing rate temporarily increases, what could this mean? Does it imply that the feature to which the neuron is tuned becomes bound to, or grouped with, other features? The alternative would be that the increase arises because another feature is shown to which the neuron is still better tuned, or because other conditions arose to stimulate that neuron by means other than binding. Under a pure firing-rate coding regime the significance of a change is ambiguous: it could signal binding or other processes.

Another problem with a pure firing-rate account is that firing rate, by definition, needs a particular time window to be determined. Typically neurophysiologists use a range of hundreds of milliseconds to determine firing rate, as most neurons fire at a rate roughly between 0.5 and 30 Hz. Perceptual decisions, however, can be based on very short time windows. Visual detection can be successful with stimuli as short as ~30 ms in duration (DeHaene et al., 2001). In this narrow time window for stimulus processing, the issue of firing rate is by and large reduced to the question of whether the stimulus lasted long enough to activate a neuron or not. We will dig further into mechanisms relying on spike timing below, in particular *phase coding*, of which zero-lag spike synchrony is a special case.

Altogether, binding and grouping of features remain an essential component of low-level integration and thus an important building block in constructing conscious representations. Higher-order feature detectors are likely to contribute their own share to feature integration, which is limited, however. Important resources for feature binding lie in early, pre-attentive grouping processes and in top-down influences from attentional and memory systems.

9.3 Modality Identification and Phenomenal Meaning

In addition to feature binding, our previous discussions prompt examination of another type of sensory integration, which may be placed at a higher organizational level because it reaches beyond a single sensory modality: modality identification, and hence phenomenal meaning. The sheer diversity of unique feature detectors clarifies in part why many complex, inanimate systems in nature do not qualify as being conscious. However, figure 8.4 made painfully clear that simply putting together modality-specific neurons in a pie chart of interchangeable units falls short of convincing us that such a configuration generates meaningful experience. We may respond to this in two ways. First, we may admit—with some defeatism, but perhaps also realism—that there will always be a gap between

neuron-based descriptions and subjective psychological processes. Second, we can explore more deeply how far biological neuronal networks may bring us toward fulfilling the requirements for conscious systems, and how we can break up the large “gap” into smaller valleys that might be bridged.

Recalling cardinal cells, we witnessed some fundamental problems of accounts based on feature detectors that remain unconnected, apart from their sensory afferents: combinatorial explosion, poor flexibility, low fault tolerance, and lack of capacity to generate meaning. Various mechanisms may explain how features can be bound into objects, but they do not clarify how the brain comes to know what these features *are*. Here we need to work toward a more principled approach that neither evades nor ignores the problems, especially the issue of “content.” A first principle can be derived from the tenet that the specific contents of a conscious representation will have a specific counterpart in neural activity patterns (cf. Pennartz, 2009):

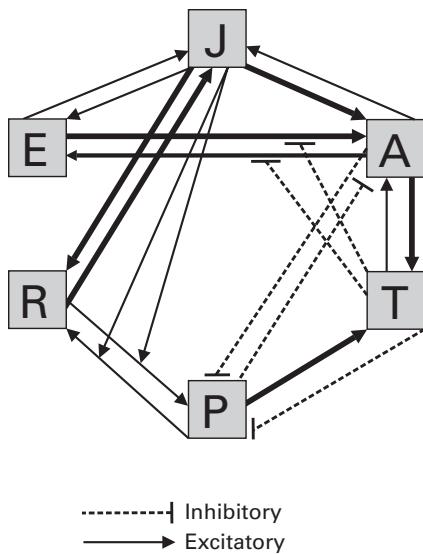
- (i) For each unique percept, the underlying neural mechanisms generate a unique configuration of activity, ensuring that two different percepts cannot be represented by a single neural state. Vice versa, identical neural configurations result in the same percept.

One and the same percept may be argued to tolerate some variation in neural configuration, as we may assume that the contributing neural assemblies have a certain robustness.² If a configuration is determined by millions of neurons, a single-cell change in activity should not matter for the percept, and it is a matter of debate how large the critical mass should be to change it. Working with channelrhodopsin-expressing neurons in mouse somatosensory cortex, Huber et al. (2008) estimated that a neuron count in the order of 61 to 297 is sufficient for influencing an animal’s behavioral report of detecting an optically induced stimulation of its somatosensory cortex.

To begin dealing with problems arising from unconnectedness, a second principle is proposed to ensure interactions between feature detectors:

- (ii) to prevent configurations of stand-alone feature detectors, neurons engaged in one and the same representation must influence aspects of each other’s neural activity.

These aspects may pertain to time-averaged firing rate properties versus properties based on precise spike timing. The importance of causally relevant interactions between neurons in constructing conscious representations has been emphasized before (e.g., Tononi, 2004; Seth, 2008), but a further exploration of the nature of these interactions is in order. At the single-cell level, spike timing itself is not a meaningful property because it should be referenced to other signals, such as EEG oscillations recorded from surrounding tissue. Evidence favoring the importance of spike timing for coding of information is currently gaining momentum (e.g., Kayser et al., 2009; Siegel et al., 2009; O’Keefe & Recce, 1993; Ahissar & Arieli, 2001; Huxter et al., 2003). By itself, this recent evidence does not yet demonstrate the use of spike phase in the coding of feature relationships, but it does

**Figure 9.5**

The naming game. The task for a neutral observer is to identify persons J, A, T, P, R, and E by observing videotapes displaying their behavioral interactions, comparing them in an ethogram specifying their correlations, anticorrelations and conditional relationships. A solid arrow represents an activity pattern in which the sender promotes a higher behavioral correlation with the receiver; a dashed line with flat ending is an inhibitory projection (sender's behavior is directed away from receiver). A projection directed at a connecting line element signifies an inhibition or excitation onto the relationship between two individuals (e.g., "person E and A like to be together unless T is around").

highlight the possibility that spike phases can be used to code other information than represented by firing rate.

Placing multimodal topologies (see figure 8.3 and 8.4) in the context of principle (ii), one may wonder how a multimodal scheme of asymmetrical relationships between sensory modalities may be effectuated by active, interconnected groups of neurons. Suppose that all sensory inputs in figure 8.4, including their relational strengths, but without any explicit label, would be fed into an *interpreter system* (cf. Baynes & Gazzaniga, 2000). This system's job will be to make a verbal (or other behaviorally expressed) judgment about the sensory inputs it receives, addressing questions such as "Which sensory modality is predominant in the current input?" or "Is the color blue part of the input?" Given the inputs in figure 8.4, how well can this system identify which modalities were active in a certain region of personal space? The proposed approach can be compared to a *naming game* (see figure 9.5).

In this game we replace each sensory modality by a person, so that we have a group of members called A, B, C, D, E, and F (for instance: vision, audition, touch, smell, taste, and vestibular sense). Beforehand we are told that one member is named John, a second one Ann, then Eric, Rita, Patricia, and Tim. You are also given some hints for solving this

puzzle, for example, “John has a close relationship with Rita” or “Patricia and Ann avoid each other” or “Eric and Ann are together unless Tim is around.” Next, you are given a stack of videotapes of these individuals, carrying labels A through F on their chest, and your task is to figure out which of these persons, A to F, can be identified as John, Ann, and so forth. From the tapes you extract statistical data telling how A’s presence is correlated or anticorrelated to B’s presence and so on. Thus you can guess with reasonable confidence that “A” corresponds to Patricia and so on.

Using its modality-specific feature detectors in computing cross-modal correlations, also a brain system may actively self-identify, or individuate, its own sensory modalities (Pennartz, 2009). Because each modality has its specific moments of peak activity, and these correlate to peak activities in some of the other, but not all, modalities, an interpreter system receiving all of these variously correlated inputs could decide “Sensory subsystem X is active and correlates to Y but is decorrelated to Z,” and therefore this configuration can be identified as a combination of touch (X) and vision (Y), devoid of smell (Z), localized at a particular spot in personal space.

But aren’t we cheating here, as the brain gets no separate “hints” as you received in playing the naming game? We implicitly assumed that “activity in X correlated to Y but decorrelated to Z” could be identified as “X = touch” and “Y = vision” and “Z = smell,” whereas this mapping is precisely what the system has to sort out itself! I argue this is not a reasoning error. If we do not wish to *explicitly* identify X with touch, our interpreter system will not be able to say “touch” when we feed in the configuration of X, Y, and Z, but it *can* still interpret the situation as “X correlated to Y but decorrelated to Z.” The claim is that this is sufficient at least for the computation of modality identification, regardless of anyone’s wish to label it as “touch” or give it another, arbitrary name. In a configuration where X correlates to Y but is decorrelated to Z, a good name to assign to X would be “what is correlated to Y but is decorrelated to Z.” Because this becomes rather verbose for an interpreter that should produce speech efficiently, a handy shorthand label will be “touch.” What matters is that modalities can be identified by this relational mechanism. The overall operation will be an active process because how our brain interprets a multimodal situation depends on *all* inputs’ being around, whether barely active at background levels or reaching peak activities. Because representations are active, constructive phenomena, simply having the anatomical connectivity in a dispositional form will not do.

At this point I am not claiming we have explained phenomenal experience; figure 9.5 merely provides a logical basis allowing neural ensembles to compute correlational or topological relationships distinguishing sensory modalities at a basic level. Analogous to the who-is-who game, we have not yet thoroughly addressed who (or what) “watches the video,” or what the nature of an interpreter is. Nonetheless, this approach can also be used to differentiate submodalities, or attributes of objects and situations in general. The interpretive process should be able to proceed quickly, which is feasible because the brain’s multimodal networks have been extensively pretrained during pre- and postnatal

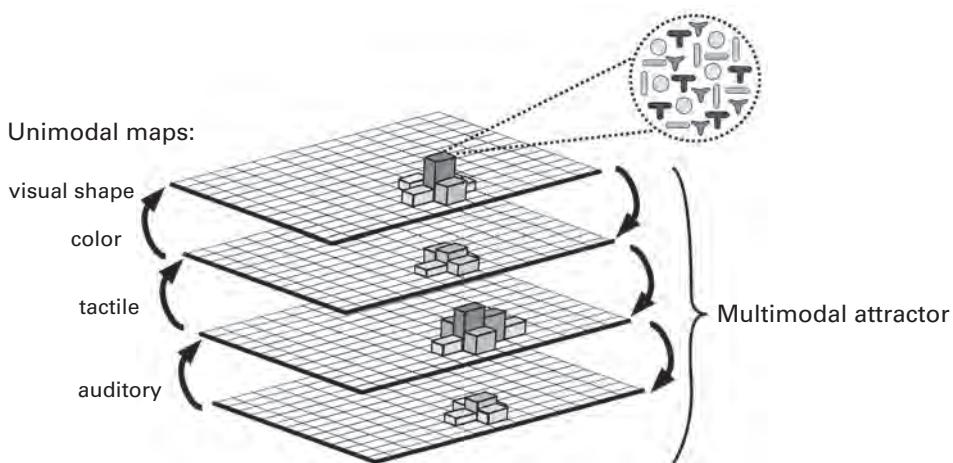
development. They harbor a synaptic connectivity that already reflects one's record of commonly experienced correlations between modalities ("multimodal congruency"; Pennartz, 2009; cf. Avillac et al., 2005; Stark et al., 2006; King, 2009).

Our next step will be to explore how whole objects and, eventually, entire multimodal situations may be coded. Here the question is how interactions between low- and higher-order feature detectors can be specified further, showing what *kind* of neural activity configurations can be applied to bring the concepts of spatial maps, feature binding, and modality identification together (Pennartz, 2009). Figure 9.6 shows a multimodal stack of sensory layers in which each layer represents a spatial map. The uppermost layer is a two-dimensional map for visual shapes (roughly corresponding to area IT in primates and fusiform gyrus in humans), the second layer a map for color (corresponding to area V4/V4 α and related, temporal areas in humans; Zeki, 2001; Conway & Tsao, 2006), the third and fourth layers maps for localized tactile and auditory inputs. This graph makes a couple of assumptions. First, for convenience we limit ourselves to two-dimensional maps, whereas in practice three-dimensional localization is at stake. Secondly, various modalities work in distinctly referenced mapping systems; in figure 9.6 maps have been artificially aligned so that a position in personal space is given by the same (X, Y)-coordinate on every map.

Looking at the visual-shape map, each location harbors a wealth of feature detectors that will signal the presence of a specific shape (e.g., a "T") on that particular location in the visual field. Imagine, for instance, a toddler playing with letter-shaped objects and trying to fit them into a cube having matching letter-shaped holes on its sides. The T-shaped object will elicit a localized "bump" of activity on the visual-shape map, and once the child touches the letter object with his fingertips, a subset of feature detectors on the matching tactile-map location will be activated. Nothing unusual so far. If the letter object is blue, this will coactivate a hub of blue-sensitive neurons on the color map, and so on for audition. The scheme can accommodate lateral connections between map locations—for instance, to enhance spatial contrast.

To satisfy principle (ii), we now connect the maps to each other topographically, so that feature detectors at an (X, Y)-location in the visual shape layer send axonal projections to cells at the corresponding location in the tactile layer and so on. A dense network of cross-modal connections arises. The underlying philosophy is that phenomenal meaning arises, in a bottom-up fashion, from the *relations* between sensory feature detectors. Importantly, the neural model system (see figure 9.6) must code spatially localized information about features and their relationships to other features *simultaneously*.

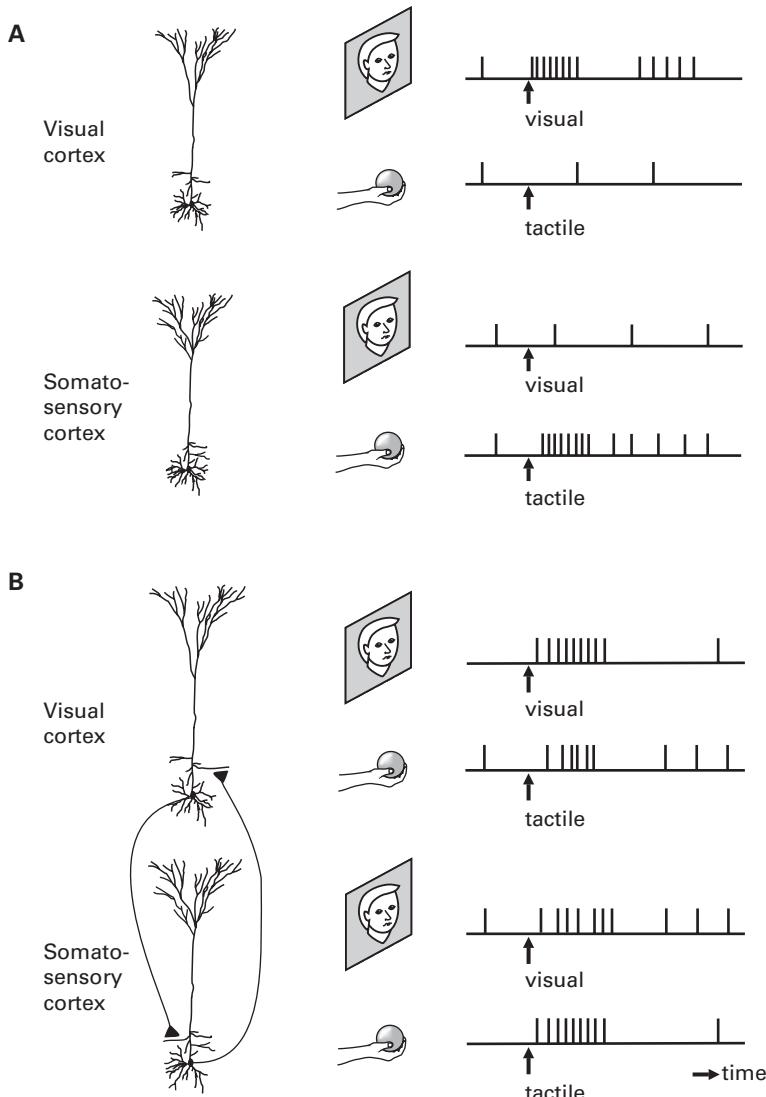
Consider what would happen to the specificity of feature detectors in each map once they are influenced by inputs from other modalities. If any detector's specificity is determined merely by its firing rate, a tactile or chromatic input influencing a visual-shape neuron would degrade its specificity (see figure 9.7). This may pose a serious problem because, in fact, single-unit recording studies do suggest that firing rate is the predominant parameter neurons use to code feature information (e.g., Oram et al., 1999; Shadlen & Movshon,

**Figure 9.6**

A dynamic associative multimodal network as a set of interconnected sensory maps. Higher-order cross-modal areas and subcorticothalamic inputs have been left out for convenience. Maps are shown for the visual, tactile, and auditory modalities. The two upper sensory maps pertain to two visual submodalities, namely, visual shape and color. Up- and downward arrows symbolize reciprocal, widespread connectivity between two successive layers, but also between layers further apart. At least part of this intermodal connectivity can arise indirectly by way of top-down influences mediated by higher-order cross-modal areas. A rectangle on any sensory topographic map denotes a small locality in the joined coordinates of a common spatial framework (which is assumed here for the sake of clarity) and contains a group of feature detectors for the relevant modality. The situation illustrated shows a spatial profile of mean firing activity for each map. Mean activity levels are enhanced for a common spatial position which is surrounded by a rapidly decaying gradient of moderate activity (dark gray corresponds to high, light gray to lower mean activity). The spatial locality commonly activated across all maps indicates the presence of an object activating all of the modalities shown. The inset for the visual-shape layer shows that the mean firing-rate enhancement for the common locality can be decomposed into high activity of feature detectors tuned to a T-shape (dark gray shading) and lower levels of activity for feature detectors less well tuned to that shape (lighter gray shading). The quasi-stable state associated with sensory coding in each particular layer is referred to as a unimodal attractor, whereas the overall state is regarded as a global-multimodal attractor. From Pennartz (2009), with permission from Elsevier.

1999; Shadlen & Newsome, 2001; Huber et al., 2008). Thus, a pure rate-coding scheme has the strong disadvantage of utilizing a single variable (firing rate) for two purposes simultaneously: to detect a specific sensory feature and to signal this feature's meaning relative to other (sub)modalities (Pennartz, 2009). Although a partial dilution of a firing-rate code might not completely destroy the system's coding capacity, it would nullify the efforts of peripheral and central nervous systems to maximize tuning specificity and sensory discrimination capacities, so this scheme is unlikely to apply in practice.

In short, the system is in need of another dimension for coding feature relationships while preserving feature specificity. Inspired by the earlier theories on neural synchrony, which relied on near-simultaneous firing, the *phase dimension* of firing—relative to a common periodic event such as a mass oscillation—was proposed to serve this purpose (Pennartz, 2009). Because of their higher frequency and local specificity, gamma- and

**Figure 9.7**

Degeneration of firing-rate specificity by multimodal mixing. (A) Configuration with segregated processing of visual and tactile information. Visual stimulus presentation (face) results in excitation of a pyramidal cell in the visual cortex, as apparent from an increase of firing rate above the background level. Onset of the tactile stimulus (ball in hand) does not alter its basal firing rate. The same applies to the firing-rate specificity of tactile processing in the somatosensory cortex (lower panel of A); no multimodal interactions occur. (B) If feature detectors in the visual and somatosensory systems influence each other directly or indirectly, as symbolized by the reciprocal connections between the two cells, the tactile stimulus will be able to alter the firing rate of the visual cortical cell and the visual stimulus will modulate the somatosensory cortical cell. The strength of the effect depends on the net weight of the reciprocal connections. This type of multimodal cross talk will result in a degeneration of the specificity of feature detectors as far as depending on firing rate.

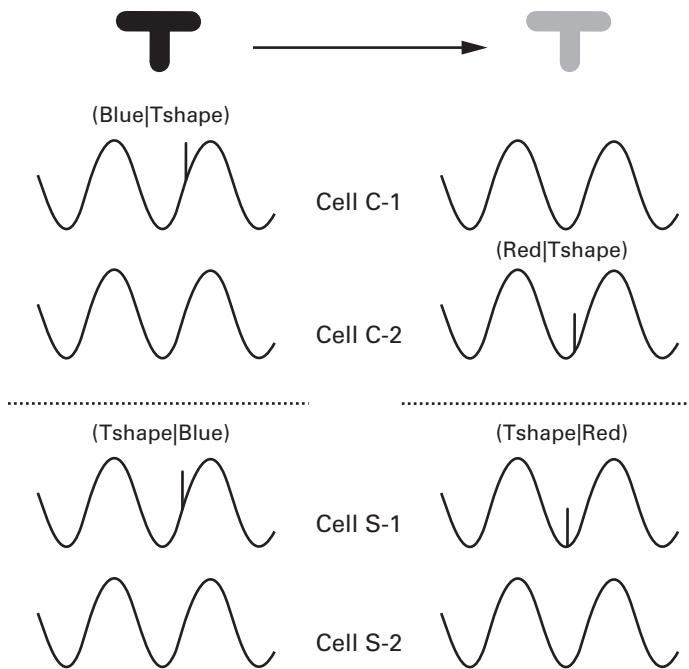


Figure 9.8 (plate 11)

Phase coding of multimodal relationships with retention of firing-rate specificity. Upper left: The color-sensitive cell C-1 is tuned to blue and thus fires in response to blue, whereas cell C-2 is tuned to red and will remain unresponsive because the presented T-shape is colored blue. Lower left: Likewise, shape-sensitive cell S-1 is tuned to a T-form and thus fires upon presentation of the blue T-shape, whereas S-2, tuned to circles, remains silent. Upper right: When the color of the T is changed from blue to red, the red-sensitive cell C-2 will now be activated whereas C-1 will drop silent. Lower right: In the shape layer, the T-sensitive cell S-1 will remain active as the shape of the object is maintained, while S-2 remains inactive. The crucial detail is that the phase of firing of the T-responsive cell changes when the stimulus switches from blue to red while retaining the same shape. From Pennartz (2009), with permission from Elsevier.

perhaps beta-band activity may prove useful although involvement of lower frequency bands (theta or alpha) is not excluded.

Figure 9.8 (plate 11) considers a phase coding scheme applying this idea to the combination of a T-shape and color blue. Visual processing leads to activation of “T” and “blue” detectors in the two upper layers of the stack (figure 9.6). On a wide time scale, the activation is expressed as a global enhancement of firing rate, but on a finer scale the spikes are fired at specific phases of an oscillatory cycle. For instance, when the color of the T-shape is changed from blue to red, the constancy of shape coding is guaranteed by the T-sensitive cell S-1 (but not S-2) remaining active, whereas the firing of the same cell undergoes a phase shift relative to the oscillation cycle to incorporate the T-shape’s change of color.

It is precisely the latter effect encoding the altered association between shape and color. The coding of the T-shape is thus identified by the number of spikes each shape-selective

cell fires but is also predicated on the color condition copresented with shape: hence the notation “Tshape | Blue” (T-shape under the condition of blue). Compared to rate coding, the combination of rate and phase coding has the advantage of coding two essential stimulus variables (presence of a feature vs. its relationship to other features) by different neural variables (firing rate and phase). This aspect of the current theory—which awaits experimental testing—holds that meaningful perceptual content arises only when engaging both coding mechanisms. Simply generating a “discriminative state” by activating feature-detecting neurons is not enough for integrated perception.

Lesions of the “essential node” MT-V5 cause akinetopsia but leave perception of other submodalities such as color and shape intact. Is this compatible with the assumption that neural activity in the corresponding areas (V4–V4 α and other regions of IT) is subject to influences from motion inputs? When area MT–V5 is lesioned, motion coding by neurons in this area is lost, but the additional prediction here is that MT’s effect on other sensory areas disappears. Firing activity in these other areas is not affected by motion input any longer, but they retain their own submodality-specific coding, and they can still convey this information to *their* target areas (except MT/V5). As long as enough target areas are left intact, sufficient space for relational coding by phase shifting is left, and so the loss of a single area will not greatly affect perception via other (sub)modalities. In conclusion, the traditional way of thinking about sensory lesion evidence has been that object properties are all essentially independently and separately coded in different brain areas. However, the available evidence is equally compatible with the idea that a feature change in one modality changes aspects of representation in another modality (cf. Zhou & Fuster, 2000; Shimojo & Shams, 2001; Macaluso et al., 2002; Fu et al., 2003; Macaluso & Driver 2005; Schroeder & Lakatos, 2008; Iurilli et al., 2012).

The importance of having background activity is further motivated by the “consciousness in a bottle” problem (cf. Block, 2005). The question addressed is how far we can strip a neural circuit down to its bare necessities before consciousness is lost. If all cortical areas in the primate visual system were lesioned except for V1, V2, V3, and MT/V5, with intact feedforward and feedback connections, would this reduced system still support conscious motion vision? Following the current theory, we translate this problem to a situation where all modalities except one are completely inactive. There would be no way for the system to relate the information from this single modality to other modalities and thus identify the modality activated by the stimulus. For instance, if a visual motion stimulus is presented, area MT/V5 will become active, but high-level sensory cortical areas of other modalities would remain silent. This will effectively reduce the system to a single, unconnected set of feature detectors (see figure 4.10), except for the connections within the remaining visual structures. The current theory predicts that no conscious *vision* will exist, because the remaining processing can no longer be individuated as “visual” as the required references are lacking. In conclusion, “consciousness in a bottle” is not a feasible scenario.³

9.4 Neural Mechanisms of Cognitive Interpretation

A key requirement for conscious representation is to interpret inputs unambiguously, not only phenomenally (in a perceptual–sensory sense), but also cognitively, such as when switching between Jastrow’s duck and rabbit. Cognitive interpretation is related to, but arguably of a higher order than, sensory interpretation. How can cognitive interpretations arise in neural systems, and to what extent are they similar to lower-level processes of phenomenal interpretation?

In advance of cognitive interpretation, lower-level visual information will be propagated feedforward to higher brain areas involved in the semantic classification of inputs. The collective activity of feature detectors in V1, V2 (etc.) representing the visual inputs of figure 4.5 will, by itself, not suffice to “see” the input pattern *as* a duck or a rabbit. To see a duck, this activity pattern needs to activate cell populations higher up in the visual-mnemonic hierarchy, which will *generally* fire when a duck is perceived: these cells code the category “duck” whenever an input sufficiently representing a duck reaches them. This activation might be achieved by the visual input as rendered in figure 4.5 but also by other visual inputs such as body movement patterns or “quacking” sounds of a duck. Once the “duck” category cells are activated in this bottom-up manner, their own activity will be propagated recurrently to the lower-level assemblies conveying the input (cf. Deco & Rolls, 2004; Camprodon et al., 2010; Wyatte et al., 2012).

Is this theoretical scenario physiologically realistic? Current evidence attributes category-coding functions to neural assemblies in the prefrontal cortex, inferior and medial temporal lobe (MTL; Freedman et al., 2003; Meyers et al., 2008; Quiroga, 2012). In chapter 3 we encountered the cognitive disorder *semantic dementia*, marked not by a loss of concrete daily-life memory but by the inability to retrieve and use general, decontextualized concepts to describe facts (Hodges & Patterson, 2007). Such concepts are exactly the categorical information we are looking for—the materials used to cognitively interpret sensory inputs. Specific semantic deficits are intimately associated with damage to the human medial temporal lobe, more specifically the anteromedial and inferolateral temporal cortex (Noppeney et al., 2007; Binney et al., 2010).

Supportive evidence comes from electrophysiological recording studies in awake patients undergoing electrode implantations to localize the focus of epileptic seizures. While the neurosurgeon is lowering electrodes in search of the epileptic “hot spot,” it is possible to record single-cell firing patterns in, for instance, the medial temporal lobe. Meanwhile, the patient is asked to view pictures of well-known people or objects. In a landmark study, Rodrigo Quijan Quiroga et al. (2005) reported specific MTL cell responses to faces of celebrities, such as Jennifer Aniston and Bill Clinton. When pictures of Jennifer Aniston were repeatedly shown, the cell reliably fired a burst of spikes each time the picture appeared and responded less or not at all to other faces. That there is a reasonable chance of finding a cell that responds to a specific face during a single electrode

descent—neurons respond, on average, to about 2–3% of the stimuli presented in a set—demonstrates there must be many cells responding to the actress but that go unrecorded; otherwise the chance of hitting that single, unique Aniston cell would be almost infinitesimally small. The number of cells specifically tuned to a well-known face was estimated at less than 1 million neurons per medial temporal lobe (which contains about 10^9 cells; Quiroga, 2012)—which, incidentally, supports the brain’s use of a sparse population code for personal identity.

However, why should we think of these person-specific cells as coding a *category*, and not the complex sensory pattern that was shown, as in the hypothesis of a grandmother cell? In other words, is a cell responding selectively to a person because of the specific sensory features conveyed by that picture or because the response corresponds to the *recognition* of the actress, which would happen even if aspects of the presentation were to change? In their 2005 study, Quiroga and colleagues also found a cell responding to Halle Berry, but the novel finding was that this cell kept on firing even when Berry was disguised as Catwoman. Different visual appearances of the actress all elicited a reproducible response. Strikingly, the cell also reacted to the written or pronounced name “Halle Berry.” Apparently, inputs from multiple modalities were capable of activating the cell, as long as the inputs were all unambiguously coupled to “Halle Berry” (see figure 9.9, plate 12). Altogether these MTL cells should not be classified as “sensory” cells but as “categorizing” units coding an abstract representation or “concept” (Quiroga, 2012).

It is not far-fetched to pose the existence of categorizing cells responding to more abstract categories of objects, such as chairs, ducks, grandmothers, houses, or animate things versus inanimate objects. Even before Quiroga et al. published their person- or concept-related cells in 2005, category-specific neurons were reported in human hippocampus, amygdala, and entorhinal cortex (Kreiman et al., 2000), and occasionally patients have been identified with category-specific lesions in their brain. A group of patients who had suffered from herpes simplex encephalitis, for instance, was unable to identify living things or foods but could classify inanimate objects correctly (Warrington & Shallice, 1984). How category-specific cell populations are formed during early and adult life is not known, but the selectively responsive cells that Quiroga recorded were all concerned with personalities the patient was familiar with, even when the subject had only recently met the person.

Figure 9.10 lays out a rough metaphor for a neural implementation of cognitive interpretation. In figure 9.10A, two mirrors are facing each other, reflecting patterns of light into an endlessly regressing and diminishing sequence of images. This metaphor is different from what we are aiming at, because in the brain we need at least one “visual input” station and at least one higher level to represent a nonvisual (categorical) feedback. Next, imagine what will happen if the right mirror transmits a different image than the left mirror projects back. For instance, a Jennifer Aniston image is propagated from the right, but the left mirror projects back Kate Winslet. If we peek into the system to the left or right, we only see the actress being projected out of each respective mirror, although the two-mirror

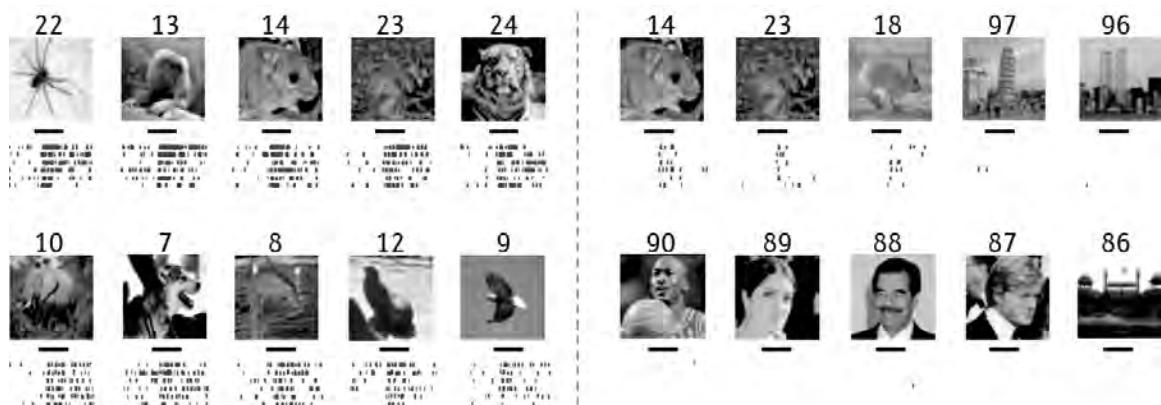


Figure 9.9 (plate 12)

Single-cell recordings from human epileptic patients. Recordings were made using intracranial depth electrodes implanted in brain areas such as the hippocampal formation and amygdala. The left part of the figure pertains to an amygdala neuron that responded with an increase in firing rate to pictures of animals, but not to other kinds of stimuli (e.g., faces or places; these classes are not shown). The right half is from another amygdala neuron, which fired only in response to the mouse, squirrel, and rabbit, out of 97 pictures shown. The black bar below each picture shows the time window of stimulus presentation (1 second). The display below each bar shows, for each row, the moment the cell fired relative to the stimulus, with each vertical blue tick representing one spike. Subsequent rows represent consecutive stimulus presentations (trials). Note how the right-hand neuron not only fires more selectively to the three animal pictures but also hardly fires at all to pictures of buildings and human faces. The left neuron was putatively classified as an interneuron, the right neuron as a pyramidal cell. Such cells are thought to code abstract concepts, such as “animal” or “rodent” in a sparse and explicit manner. From Quijan Quiroga (2012). Reprinted by permission from Macmillan Publishers Ltd.

system as a whole contains information about both actresses. As yet, there is no integration of the two images because the projections do not affect each other.

At the next step, the right mirror projects visual input to the left one, but now we suppose that the left mirror projects back information that is of a different kind than coming from the right. If the right mirror conveys low-level, detailed visual information such as on orientation and length of line elements, the left mirror is now influenced by this input but projects back information about larger, complex shapes. In turn, this feedback pattern reshapes the patterning on the right mirror, highlighting particular low-level features such as boundaries and resulting in a *modified* image projected back to the left mirror. These back-and-forth projections will dynamically evolve over time.

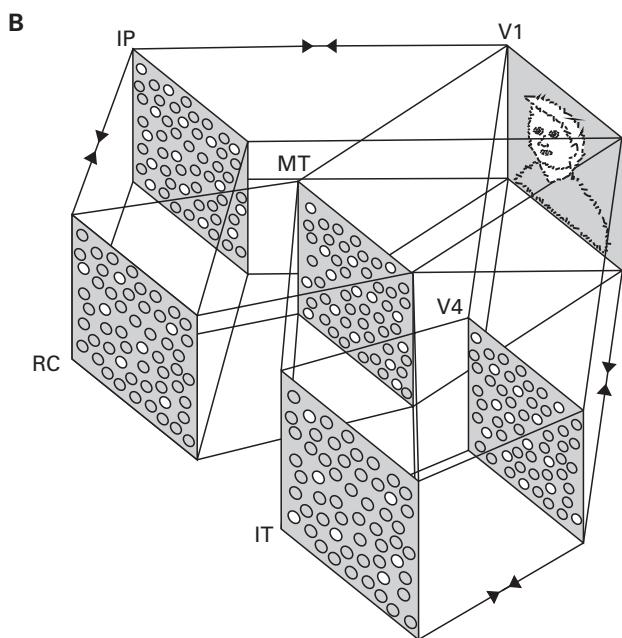
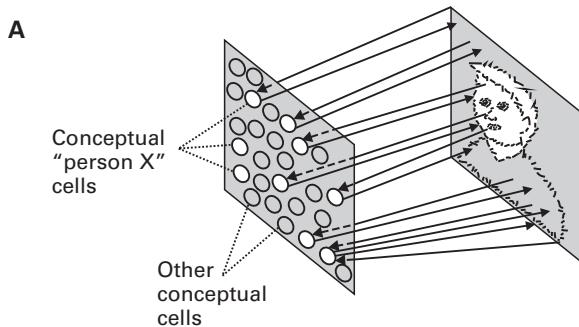
For cognitive interpretation, we adopt this scheme but let the left mirror project *categorical* information back to the visual-input stage. The low-level coding elements elicit a semantic feedback from the left mirror, which modifies processing on the right side, and so on iteratively. Thus the left mirror projects *mnemonic* knowledge, not sensory input, back onto the first mirror. As the thought experiment progresses, it becomes harder to *imagine* what the reciprocal interactions “look like.” This is because you are commanding yourself to imagine processes that cannot be “sensorily” imagined and are of a more abstract nature.

If I imagine low-level “Charlie Chaplin” information being projected from the right, I am already making the mistake of imagining his *picture* to be there—but at this first stage no comedian is present, no personal identity, just a bunch of neurons responding to low-level visual elements. Imagining the semantic rendering of Chaplin on the left is also impossible, because I can only imagine a face, written or spoken name, or some other sensory attribute associated with the great comedian. Imagination is relatively useless when it comes to understanding conscious representation, yet we rely on it constantly when trying to get hold of it.

This echoing-mirrors scheme of cognitive interpretation is prelinguistic, but it is instructive to compare the process to the construction of basic propositions. If we perceive a cat, for instance, the sensory input serves the role of the demonstrative pronoun “this,” whereas the feedback from semantic memory represents the concept “cat”: *This* (i.e., the current sensory input) *is a cat* (i.e., a specific category). The verb of incomplete predication “*is*” marks the interaction tying sensory input to interpretation. The sensory input, coded for instance by V1 cells, does not mean anything *by itself*—it indicates “*there is an input here*” but does not specify what the input *is* to begin with. It is just an input without interpretation. At the other end, high-level cells in associative cortex are taken to signal the category “*cat*,” but at that level *alone* it is unclear to which input it belongs. The concept is “*cat*,” but the input details are lacking. To see a specific cat, both areas are needed, underscoring why perception is a systems property. All of this proceeds in the sensory–perceptual domain, without involvement of actual verbal propositions.

Figure 9.10

(A) A lower and higher sensory cortical area represented as an opposing-mirror system. Dashed line elements in the right-hand area represent visual cortical cells selective for oriented line elements in small receptive fields. The overall population pattern represents a face, but the input has not yet been identified as a face at this lower stage. Leftward arrows transmit line element inputs to the left-hand area that integrates these inputs into larger receptive fields. The left-hand cells do not represent physical properties of the inputs such as orientation but rather the concept represented by the visual input, such as person X. The input activates only a subset of the conceptual cells (white ellipses), selectively responding to person X. The conceptual cells project their coding pattern back to the right-hand layer to affect processing of the input. To see person X, it is proposed that both bottom-up and top-down interactions are necessary: the lower area provides the input (“this”) that is interpreted or categorized by the higher area (“person X”). (B) The input at V1 is now broadcast to higher-order visual areas such as the middle temporal area (MT), V4, and intraparietal cortex (IP) and henceforth to yet higher areas such as the inferotemporal cortex (IT) and rhinal cortices (RC; including entorhinal and perirhinal areas, projecting upstream toward the hippocampus, not shown). Not all intermediate areas have been shown as part of these routes (see figure 6.3, plate 3, for a more complete mapping). The reciprocal, echoing interactions are symbolized by colliding arrows (not shown for all map-to-map connections). The input activates subsets of neurons in MT, IP, and V4 and even sparser subsets in IT and RC. Recognition of complex object and situational information in IT, RC, and hippocampus results in feedback to higher visual areas where, for example, motion- and color-coding patterns of activity are modified as part of the recognition process. Conscious representation cannot be localized to any particular node of the system as each node makes its own perceptual–cognitive contribution. The scheme leaves out many relevant visual and memory-related areas as well as other sensory modalities and multisensory areas such as posterior parietal cortex.



Many types of recurrent and lateral feedback to lower cortical input stages are not included in this two-mirror model. For vision, these tentatively correspond to V1, V2, V3, and ventral posterior occipital cortex. Another simplification is that processing ends at the left mirror, roughly equivalent to the MTL categorical–semantic system, as also parietal and prefrontal cortex exert top-down control over which categories may be selected for further processing and how attention is directed. These omissions are not irrelevant but may be temporarily tolerated for the sake of clarity.

A more essential expansion of the scheme is to add *many* left mirrors in parallel now, as modally specific input from the right projects to many higher sensory and categorical areas, providing multiple kinds of multimodal and interpretive feedback. As V1 information is sent to extrastriate areas and higher associative areas such as IT and perirhinal cortex, each of these areas will reflect their specific information back to lower visual areas—but will also interact with each other as sensory–cognitive analysis proceeds. For example, once a specific object has been isolated from the background and classified via IT and related MTL areas, the spatial relationships between situational elements can be analyzed by way of interactions between visual, vestibular, motor, parietal, and parahippocampal areas (Goodale & Milner, 1992; Snyder et al., 1998; Kornblith et al., 2013; Dilks et al., 2013).

Grossberg's adaptive resonance theory (ART) already made use of bidirectionally connected layers in network models simulating sensory processing and object classification (Grossberg, 2007, 2013). Briefly, ART models make use of a lower layer of feature detectors which code an input pattern; an upper layer that adaptively processes the input contains category-selective cells that will be activated given an appropriate input. Aided by a vigilance and arousal system, these bottom-up and top-down interactions lead to a recognition process identifying matches or mismatches, where a match is accompanied by a synchronous resonant state embodying an attentional focus and supporting fast learning. While ART models have significant merit in the faculties of learning and recognition, it is less than clear why, or how, resonance between the two layers would give rise to consciousness. For instance, ART predicts that “all conscious states are resonant states,” but does not predict that “all resonant states are conscious states” (Grossberg, 2013, p. 6). The account, however, fails to clarify what is missing about some resonant states that would otherwise render them conscious.

A major difference between ART and the current framework is that the latter focuses on how a neural system specifies or individuates an object's heteromodal qualities relative to other properties and other objects in the situation. When these qualities are understood broadly and include conceptual aspects—in addition to aspects dealing with modality identification—a functional multiplicity at higher-order analysis stages is called for (see figure 9.10B), contrasting to the core two-layer architecture of typical ART models. If we take, for instance, a single, colored line element in a Mondriaan painting, feature detectors in the lower visual system will vigorously respond to it, and Grossberg's resonance between the lower and higher ART layer results in the object's being classified as a

particular kind of line element. Whereas ART models can provide for context-sensitive match states, their description does not substantiate how the surroundings of the line element are coded themselves, and are corepresented in proximity to this element to represent a scene. A system for conscious representation should do so. The Mondriaan painting may show, for instance, a blue line against a white background, and further red and yellow lines in parallel or orthogonal orientations in the periphery. Referring to the multimodal stack of connected maps (see figure 9.6), all elements represented in a conscious “summary” of the situation will spontaneously elicit localized bumps on the maps of the relevant modalities. Additionally, semantic-interpretive systems interact with these sensory representations to enable recognition both of single elements in isolation and of the broader Mondriaan configuration.

The scheme of multiple systems continuously echoing module-specific information back to lower stages and to each other does not treat feedback from higher “visual” areas such as MT fundamentally differently than feedback from still higher “categorical” areas such as the anterior temporal lobe. Following chapter 8, we do not regard sensory interpretation as being radically different from cognitive interpretation, despite graded differences in complexity and content.

Having a graded scale between sensory and cognitive interpretation also has consequences for our thinking about the distinction between phenomenal versus associative meaning. “Tomato” derives its *associative* meaning from its linkage to other items in its semantic network such as “vegetable” and “spaghetti sauce” (see figure 5.2). So far, we have treated phenomenal meaning quite differently, maintaining that a tomato’s color is not dependent on the retrieval of items associated with “red.” Color seems to appear self-evidently to us, and there is good reason to believe that we do not need to retrieve semantic knowledge to interpret a color as “red.” For example, patients with semantic dementia have no problem seeing “red” although they may have trouble using the category of “red” to describe objects in daily conversation. However, the “Rubik’s Cube” (see figure 8.5, plate 9) showed that perceived color does depend on information presented via the surrounding context, supporting the conjecture that phenomenal meaning also hinges on interactions between primary cortical sensory input and other, intermediate-level brain areas such as V4/V4 α , but not the MTL network for classification and recognition. For visual shape perception, it is proposed that interactions of the lower (V1, V2, V3, and ventral posterior occipital cortex) with higher-level visual areas (e.g., V4, ventral occipitotemporal area, IT) are central, but also with parietal areas 5 and 7, premotor areas, and area SII in relation to cross-modal matching of shape and grasp.

A second implication is that a conscious representation is not located within any specific node of the entire system. We might find neural correlates of consciousness at multiple anatomic locations, but a more interesting quest is to identify the components of representation those locations mediate. If we record a Barack Obama cell in the hippocampus, for instance, this may crudely correlate with a conscious percept of the president

(Rees et al., 2002), but then we need to specify whether the cell’s firing correlates with a feeling of familiarity, with the semantic concept of “presidentiality,” or with particular sensory features of the picture shown. If, in some experimental paradigm, a neuron’s activity does *not* consistently correlate to consciousness over time (for instance, because it shows utrocular discrimination; see chapter 7), it could still contribute to conscious representation of a specific sort. Consider the difference we subjectively recognize when we are seeing a photo of Barack Obama versus visually imagining him: a lower visual cortical neuron that is activated by showing the photo may well remain inactive during visual imagery and might thus be disqualified as coding a “neural correlate of consciousness,” but there is still a subjective and neural distinction to be made between perception and imagination, which is determined, at least in part, by input from V1 and subsequent areas (Kosslyn et al., 2001; Daselaar, Porat, et al., 2010). Lower visual areas may contribute to perception as a class of conscious representation, but not to imagery. This perceptual representation is not at any single place within the system, in line with consciousness’s being a systems process.

A significant feature of figure 9.10 is the dynamic nature of feedforward and feedback interactions. A full situational interpretation takes time to develop. We can think of the consecutive right-to-left and left-to-right projections as an iterative process continuously updating representations in all modules. A stream-of-consciousness situation will be marked by continuous updating, but a relatively constant visual scene may drive the system to converge to a (temporarily) stable state. In case of emerging percepts, the dynamic, iterative interactions are likely to evolve sparsely and globally asynchronously, in agreement with the desynchronized nature of the wakeful EEG and sparingly distributed ensemble firing. Nonetheless, this firing may be organized in locally effective gamma cycles. Miconi and VanRullen (2010) proposed to deploy gamma oscillations for organizing “perceptual cycles,” but their scheme hypothesized that successive cycles represent different objects in a visual scene, deviating from the idea of full-scene emergence across cycles proposed here.

Illusions provide interesting case studies against which models of representation can be tested. In the current scheme, a Kanizsa triangle (an illusion evoked by three black “Pac-Man” shapes having a collinearity of their edges; see figure 4.6) will elicit low-level visual cortex activation patterns signaling line-stopped elements at multiple locations in the visual field. At the higher level of IT and related areas, the global configuration of line elements elicits neural activity selective for triangular shapes and the category of triangles. Continued viewing of Kanizsa’s figure leads to repeated, “valid” activations of neurons at both levels of processing, resulting in a clash between low-level input and high-level interpretation. Briefly, there will be representational shuttling between a fully closed triangle and the three “Pac-Man” mouths, before the brain settles into a more stabilized state, where low-level, detailed input presides over high-level constructs. When viewed in probabilistic terms, this account is compatible with the concept of the brain coding an internal, Bayesian

model attempting to infer causes of sensory inputs it receives (Rao & Ballard, 1999; Lee & Mumford, 2003; Lau, 2008; Bastos et al., 2012) although it also emphasizes the prominence of low-level details sometimes violating the model's predictions.

Similarly we can think of the duck–rabbit as a special sort of “illusion,” but here the conflict arises at a higher, “cognitive” level, involving competition between semantic representations. Let me hasten to add we have not even begun to flesh out the many kinds of interpretation used to construct or influence conscious representations. For example, according to Gibson’s (1977) notion of “affordance,” we may think of sensory inputs as eliciting *motor* representations instructing us how to approach and manipulate the objects inferred from these inputs: doorknobs require a different motor approach than hammers.

9.5 Neural Mechanisms for Perceptual Stability

In the face of the “dynamism” postulated for perceptual–cognitive interpretations, how is it that our brains can stabilize percepts of, for instance, ambiguous figures and illusions? Earlier on, attractor networks caught our attention as a neural substrate for representational stability, and indeed O’Brien and Opie (1999) and others (Haken, 2006) proposed these networks as a central element in their hypotheses on consciousness. However, classical attractor networks will not do because of the modality identification problem and a lack of spatial representational capacity. Moreover, once classic networks have entered a basin of attraction, they have great trouble escaping from it—whereas conscious systems make swift transitions to novel states. Therefore we need to examine how a stacked multimodal network (see figure 9.6), now expanded with semantic–cognitive modules (see figure 9.10), may be endowed with stability.

When we first zoom in on a unimodal map, such as for visual shape (see figure 9.6), an attractor state at this level can be conceived as a “bump” of neural activity that can move across the map. Within that hillock, at a more fine-grained level, neural activity is chiefly made up of activity increments in the neuronal subpopulation tuned to a feature of the object or situation that is currently presented (see figure 9.6, inset). Given constant input, the hillock will be stationary, both in terms of map location and feature detectors activated at that location. A robust change in input will shift the bump to another location on the map or activate different feature detectors at the same location. This behavior is characteristic of a *continuous* attractor model (Ermentrout & Cowan, 1979; Tsodyks, 1999; Wu & Amari, 2005; cf. Haken, 2006).

We next apply this idea to a stack of multimodal layers. Because unimodal maps can sustain their own attractor states, and connect to each other, the overall network is also proposed to assume an attractor state, but now of a global, multimodal nature. This assumption is motivated to achieve overall stability of perception, not leaving some aspects of the situation in a wobbly condition while others have already settled (cf. Hoshino et al., 2003). This was labeled a “dynamic associative network” in Pennartz (2009) and should be extended

to include the cognitive counterparts discussed above. Given a stable activity bump in lower-level sensory maps, a higher-order semantic module may switch between two attractor states allowed by this input (e.g., an ambiguous picture such as the duck–rabbit). This switching is regulated by top-down control from frontal cortical areas and requires energy because of the resistance of “semantic” coding states to leave their basin of attraction. A similar explanation may hold for alternating perceptual states in binocular rivalry and some visual illusions, noting that the predominance of top-down versus bottom-up control may differ per case. The robustness and resilience of representations is most salient at the level of cognitive interpretation, but our examples of illusory depth (see figure 8.2) and color perception (see figure 8.5, plate 9) suggest that stability is an important principle as well in intermediate-level areas such as V3, V4-V4 α , and MT (DeYoe & Van Essen, 1988). As eye blinks and saccades cause momentary changes in V1 activity, and firing activity is not maintained here when visual input is interrupted, attractor properties at this low level may be either absent, or present but not needed to support perceptual stability.

Earlier on, I proposed that firing rate is used to code feature specificity and that spike phases are important for coding relational aspects, providing a basis for meaning. Therefore, attractors at higher sensory and cognitive levels are predicted to be robustly structured in the phase domain as well. Simulations support the biophysical feasibility of such structuring (e.g., Martinez, 2005). A switch in the categorization of an object will not only be accompanied by a shift to a different attractor state in the corresponding semantic modules but should—by virtue of principle (ii)—also tip over connected attractors in other brain areas to assume a different state, specifically in the phase domain. A category switch will thus not only be notable because of strong firing-rate changes in the MTL but also by more subtle spike phase changes relative to, for example, beta or gamma oscillations of extrastriate visual cells tuned to features of the categorized object. It is feasible to test this prediction given the evidence for phase-specific spiking activity in visual and somatosensory cortex (König et al., 1995; Ray et al., 2008; Vinck et al., 2010).

9.6 Summary

The main take-home message from this chapter is that no central readout mechanism upon which all multimodal inputs converge is needed to solve how we come to experience sensory modalities as being qualitatively different from one another. Installing an interpreter system may be handy to attach verbal “labels” to sensory input, but this does not solve the problem of how phenomenal meaning arises. Instead, the correlative structure in multimodal networks is proposed to provide a basis for addressing this matter, while we should as yet be careful about claims with respect to explaining the subjective, qualitative nature of sensations—explaining why things feel the way they do. Dual coding of feature presence by firing rate, and feature relationships by firing phase, offers a promising scheme to realize this basis neurophysiologically. In a similar vein, cognitive interpretation

is suggested to depend on the iterative interactions between low-level sensory and high-level mnemonic systems. Combining sensory and cognitive interpretation, global percepts may achieve temporary stability by dynamic attractors in coupled multisensory and mnemonic systems.

Before we move on, let us take a step back and survey where the current theory has brought us so far. In chapters 4 and 5, consciousness theories based on classic neural networks and information theory were criticized as being subject to the “panpsychic trap.” Are we now safely steering away from this trap? The articulation of key requirements for conscious representation makes it much more difficult, if not impossible, for inanimate systems in nature to qualify as conscious. The diversity and richness of representations, the capacity to coordinate, align, bind, and self-identify elements as meaningful percepts in context—aided by mechanisms for stabilization, iterative sensory–cognitive “mirroring” and dual phase- and rate-coding—all testify how versatile and representationally powerful the brain is as compared to, for example, cyclones. Hypothetical as the mechanistic implementation of these requirements may be, they also make more explicit how sophisticated and specialized the neural machinery must be to accomplish all of the identified subtasks, such as striking the right balance between dynamic, bidirectional interactions and stabilization of percepts.

10

Levels of Representational Organization

When the low air plane skims over the water,
When the low air teaches us humility,
When the low sky is gray as slate,
When the low sky is sallow as boulder clay,
When the north wind quarters the plain,
When the north wind steals our breath,
Then my land cracks, my flat land

—Jacques Brel, *My flat land* (1962)

10.1 Why Distinguish Different Levels?

Reviewing our attempts so far to plow through the minefield of consciousness research, a skeptic may still wonder why all that firing-rate and phase-coding business would actually lead to anything like the phenomena we see and feel consciously. After all, haven't we just been discussing spike trains, neural ensembles, and brain areas that send electrical signals back and forth? Where does a phenomenal sensation like the taste of an orange emerge? In short, we are still facing "the explanatory gap" (Levine, 1983). Extending work in the previous chapters, we will explore here whether this problem can be approached via the concept that neural and phenomenal phenomena are situated at different representational levels. Arguing that this approach is fruitful, although limited by the nature of our epistemic capacities, I will continue with asking how many levels can be distinguished, what their nature is, and how the relationships between levels should be understood. If we distinguish different levels of representation, are their relationships considered "causal," "equivalent," or of a yet different nature?

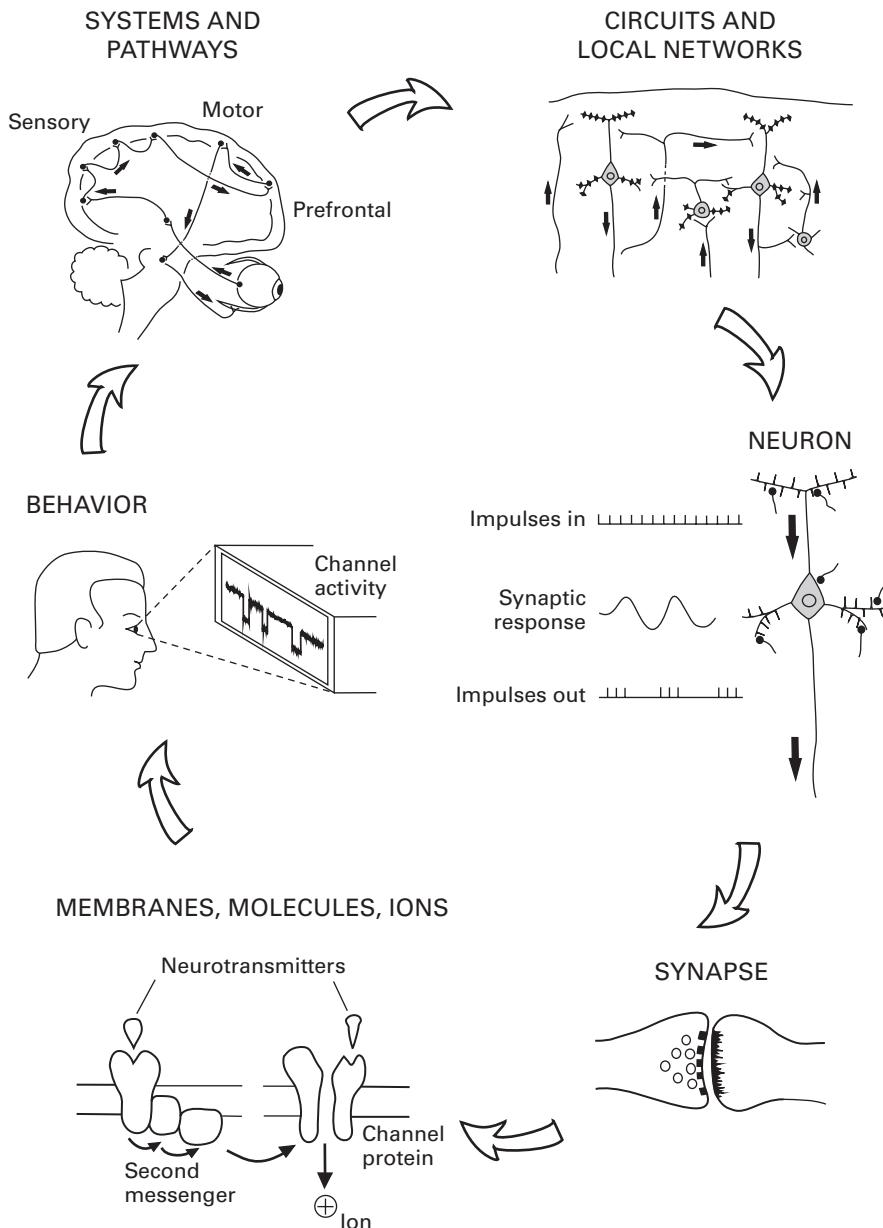
A key question is whether collective activity patterns of neurons in a given brain area can be conceptualized as "symbols" that are "read out" or interpreted, either by other brain structures or in yet a different sense. A further issue that has been burning in the background since we discussed the labeled-lines hypothesis is whether neural

activity configurations are “grounded” by sensory input or are essentially arbitrary—and how problematic this might be for generating meaningful representations.

Let us first examine whether we may get a better handle on the explanatory gap by considering different levels of neural and perceptual–cognitive organization. Classically, brain function has been defined across different *aggregate* levels of organization, such as molecules, subcellular structures, neurons, circuits, and networks, all the way up to systems of larger brain structures and behavior (see figure 10.1; Churchland & Sejnowski, 1992). Although useful for neuroscience, this distinction in levels *per se* does not illuminate at which level correlates of cognitive processing are to be situated, nor how consciousness would emerge at any level.

David Marr (1982) adopted a different view on multilevel explanation of the mind–brain, distinguishing the levels of (1) hardware implementation, (2) representation and algorithm, and (3) computational theory. For instance, in vision one can describe the neural hardware used to infer object shape from patterns of illumination and surface reflectance in images. However, by itself this description will not solve the shape-from-shading problem, which can be understood and solved by a specific differential equation computing surface geometry from intensity values in the image (Horn, 1975). The third level of computational theory specifies more abstract properties of a computational approach, such as the rationale, logic, and goals of the chosen computational strategy. Marr’s threefold distinction is a conceptual one that transcends the level of physical substrate dimension or aggregate size and thus hints that cognitive processes can be understood and analyzed at different levels of organizational or computational complexity. The approach has been criticized—for instance, because it assumes a separability of the algorithmic–representational and computational levels and because it would lack general testability (but see Kitcher, 1988). Moreover, we may need to delineate more levels or sublevels of analysis to capture different types of algorithm and representation (Rumelhart & McClelland, 1985). More important in the current discussion is that even Marr’s “highest” level does not illuminate why brains would sustain qualitative experiences when performing some of these operations.¹ Let me therefore cut to the chase and argue why a multiple-levels account of mind–brain organization should be favored. First, a restricted view on purely *neural* coding levels of organization in the brain will not bring us closer to a solution (see also Marr, 1982). Secondly, neural network models suggest how higher-level properties can emerge from an organization of lower-level, neural elements, and there is no *a priori* reason why these higher-level properties would not be able to give rise to still higher-level properties, including phenomenological experience.

For the first argument, we revisit systems neuroscience experiments in which firing patterns of sensory cortical (or other) neurons are commonly described as “coding” for a particular stimulus feature, such as line orientation, motion, smell, place, or reward value. Until now, I have been using this notion of coding liberally. The “decoding,” “readout,” or interpretation of a V1 code is supposedly done by brain areas higher up in the visual

**Figure 10.1**

Brain function is characterized by several aggregate levels of organization, rather loosely defined at different scales of measurement, beginning at the molecular level of ion channels and other proteins up to the highest level of cognition and behavior mediated by connected, macroscopic brain systems. Adapted from Shepherd (1990). Note the circularity in a person looking at his own ion channel activity displayed on an oscilloscope.

cortical hierarchy. This deferral of the “meaning” issue of a neural activity pattern is commonly seen in theoretical neuroscience, explaining meaning of activity patterns in a given network node by invoking other, connected cell groups coding for items associated with that node. However, such a horizontal deferral leads to an infinite regression of the problem. Brain regions receiving spike pattern input from a key node X can be said to “read out” the information from X insofar as they convert this input, together with other converging information, into a new output, but this output is defined at the same level of organization: spike patterns, but not phenomenology. In short, restricting the problem of interpreting neural codes to the level of single neurons or network nodes will not yield an explanation for perception.

The second argument is based on extrapolation and therefore is harder to drive home. The difficulty is augmented by my earlier assertion that classic neural networks are incapable of solving fundamental problems such as modality identification. Whereas this point remains standing for classical networks, I argued that multimodal, dynamic attractor networks harbor the capacities to overcome such limitations (see chapter 9). However, if we revisit the classic networks such as Hopfield’s associative memory (see figure 4.7), we may re-appreciate their point of outstanding relevance: they demonstrate how a higher-level property (e.g., recall of a complex stimulus such as a face) emerges² from an organization of simple neural elements obeying low-level rules such as Hebbian plasticity and synaptic integration. This is, of course, no proof that higher-level properties of neural networks will produce phenomena of a yet higher complexity, but I will argue that this concept is a feasible path to be taken.

We begin to unpack the argument here and elaborate it later on. Suppose a neural pattern A recurs over and over again in your brain while it is awake and scanning your environment. Initially, you (i.e., your entire system of brain areas) do not know what A is about. Following chapter 9, we further assume your brain has a capacity to make verbal utterances, so that you will say, at some point of scanning, “Let’s call *this* blue.” This utterance seems like an arbitrary label, but your brain will apply it consistently *over time*, as more and more objects possessing the property of blueness will pass by. This consistency is derived not only from the common property of these objects emitting light of certain wavelengths but also from the relative constancy of the way feature-coding cells are wired in the brain.

So what? We make another, major leap now. Instead of labeling the perceived pattern by a *verbal* utterance, we assume the pattern can also be labeled as an *experience*, situated (or “uttered”) at a higher level of organization than that of single neurons. Phenomenal experience comes in to replace a linguistic labeling as expressed via your vocal chords. This concept is not impossible, mysterious, or transcendental. Our brains have evolved to the point where they can select, at some point in their lifetime, how to label, or interpret, a perceived property—phenomenologically, verbally, or both.

A partly appropriate metaphor, inspired by Ryle’s (1949) call to regard the brain–mind problem as a confusion of categories, is to compare the mind–brain entity to a large

department store with thousands of employees (cf. Dennett, 1983). Its organization is hierarchical, with warehouse staff, salespersons, and cashiers working at a lower level, supervising managers at an intermediate level, and a small group of directors and board members at the top. While a warehouse worker notes there is a dense physical flux of commodities in and out of the storehouse, the high-level pendant states that “sales are going well.” Similarly, neurons are indispensable for constructing the “big picture” (and, at their own level of organization, *are* the big picture), but as individuals they cannot *see* the big picture. The weakness of the analogy, I should add, is that it equates employees (who have ideas and perceive) with neurons (which do not).

Multilevel views on consciousness and cognition have been expressed before by many other theorists, such as Oppenheim and Putnam (1958), Attneave (1961), Hofstadter (1979, 1985), Churchland and Sejnowski (1992), Wimsatt (1994), and Lycan (1996). However, the elaboration of the general idea in the following sections aims to be more detailed and empirically grounded than the previous accounts, especially in distinguishing specific levels of neural organization.

10.2 How Many Levels, and Which Ones?

When trying to envisage how many levels there could be in between “neuron” and “phenomenal experience,” we immediately run into trouble. “One the one hand,” one may say, “I am having this experience of an orange orange,” and then, frowning: “but on the other hand, I have neurons in my brain that fire when my eyes are exposed to the orange. These are completely different things and I really don’t see *the* connection. If I start grouping together neurons in ensembles and even bigger structures, I can’t tell you: look, *this transition* is where things start to look *orange!*”

Trying to *imagine* this transition is the wrong way to go. Imagination can be tremendously useful for coming to grips with many scientific problems, but this aspect of consciousness is not part of those. I believe it is actually an obstacle against solving the problem. What we *can* do is to think of the various *functional and computational demands* a conscious representational system poses on the various levels of a mind–brain system. Reconsidering the “scales of neuroscientific study” (see figure 10.1), we see how the distinction of levels in neuroscience is based on increasing physical size and different physical phenomena associated with higher aggregation levels: organismal behavior *versus* vesicular release. However, when we consider the construction of conscious representations, the question becomes this: is there a *functional* rationale for distinguishing, for instance, the “maps” level from the “systems” or “networks” level?

Transcending the single-neuron level, I propose to define the next level by delineating neuronal ensembles, because their computational and dynamic properties are functionally and conceptually very different from what neurons can do on their own. An ensemble (which is a term loosely comparable to Hebb’s [1949] “assembly” but avoids particular

connotations associated with his theory) is taken to be situated in a single (sub)region of the brain (e.g., V5/MT, hippocampal area CA3) and to consist of a collection of neurons that behave in a functionally very similar way, exert a common computational function, and have strong functional connectivity but are not necessarily anatomically grouped close together within the structure. As higher-level functional entities, ensembles may function as robust lumps or hillocks in the spatially localized coding of features on sensory maps (see figure 9.6). In attractor networks, we witnessed how ensembles can represent Gestalt or whole-pattern memories in a self-stabilizing manner and can complete stored patterns from noisy or partial input (see figure 4.4). Many other computational functions can be outlined in which members of an ensemble commonly engage although it should be kept in mind that their boundaries are not universally agreed upon or well-defined (Buzsaki, 2010).

Empirical evidence for ensemble functioning has accumulated in a diverse set of brain regions such as the hippocampus, striatum, visual cortex, and sensorimotor cortex (Salzman et al., 1990; Pennartz et al., 1994; Nichols & Newsome, 2002; Johnson & Redish, 2007; Harris, 2005; McNaughton et al., 2006; Lansink et al., 2008; Pastalkova et al., 2008; Buzsaki, 2010; Truccolo et al., 2010). On their own, ensemble recordings do not reveal *per se* how functionally similar particular subsets of the recorded neurons are, how well they are connected, or whether a recorded ensemble is sharply delineated from others or only different by gradation.³ Nevertheless, within recorded samples, subsets of neurons can be grouped together based on functional criteria (Lansink et al., 2008; Peyrache et al., 2009; Ko et al., 2011). Probably the best studied example of ensemble-level functioning comes from spatial coding in hippocampal area CA1 (e.g., Wilson & McNaughton, 1993; McNaughton et al., 2006; Tanila et al., 1997). A CA1 ensemble can be defined not only by cells firing in response to a common location an animal is occupying in its environment but also by the tight temporal coordination of firing of multiple cells (Harris et al., 2003). It is up to further empirical research to determine whether a high degree of synchronous activity is characteristic of ensemble functioning in general (cf. Izhikevich, 2006) and what its representational functions may be. In chapter 9 the usefulness of spike phase was advanced for coding feature relationships, but this principle would operate between ensembles in different areas coding distinct feature dimensions. At the level of within-area ensembles, near-simultaneous firing may serve to enhance the robustness of a local representation and computation, expressed by its signaling strength toward outside structures.

This should not raise the impression that ensembles would function all the same across widely different brain areas. The anatomy of some brain regions (e.g., area CA3) may allow local ensembles to engage in collective dynamics such as attractor behavior. Other brain structures lack such stabilizing properties. In motor cortex, for instance, ensemble activity may be better characterized as dynamically coding sequential limb movements unfolding in 3-D space (e.g., Averbeck et al., 2002; Velliste et al., 2008). Here, an ensemble

is more aptly characterized as a transiently activated group of neurons similarly tuned to a feature of an upcoming movement direction, such as where a monkey's arm will be going.

As already noted, an ensemble is defined here as being restricted to a single brain structure. What functionally delimits it from higher levels of organization is that an ensemble firing pattern does not generate phenomenal meaning on its own (see also section 9.4). The firing pattern *codes*, one can say, but there is no single other area interpreting this code to suddenly yield perception. Other single areas generate their own output, which is also phenomenally meaningless *on its own*. Ensembles in multiple structures are required to communicate to enable interpretation both in unimodal and multisensory domains and in the cognitive-mnemonic domain. In addition to the arguments given in section 9.4, further support for this notion comes from the importance of context in defining the meaning of a singular element in a scene or situation (e.g., the "Rubik's Cube," figure 8.5, plate 9, and the concept of "chair," figure 5.2; see also Hofstadter, 1985, and Smolensky, 1988, for comparable observations originating from AI on symbolic and subsymbolic systems). This cross-areal communication is envisioned not simply as a convergence from different unimodal areas onto a common target area but as a *network coding* of relationships as exemplified by a back-and-forth echoing between multiple structures (see figure 9.10). For building a conscious representation we need at least one higher aggregate level than the ensemble (or within-area network) level. I suggest the next, higher level is best characterized as a network of ensembles—a network spanning several brain areas, each containing sparsely distributed ensembles or small-scale networks specifically coding different aspects of experience (see figure 10.2). Within this network, ensembles interact to communicate their patterns to each other and enable primordial but behaviorally useful interpretations. I will refer to this level as a "meta-network." It is quite possible that this level itself can be parceled further into sublevels. In the visual domain, for instance, different ensembles—specialized in coding orientations, disparity, motion, and spatial phasing—may cooperate to collectively compute stereopsis information situated at a higher functional-computational level than networks specialized in only one attribute, such as visual motion. In turn, this higher level can be argued to contribute to yet higher visual functions such as perception of distance and 3-D motion (DeYoe & Van Essen, 1988; Rose, 1995).

Do we need still higher levels to account for conscious representation? The full, multimodal richness of conscious experience was argued to transcend the level of a unimodal system, which we now consider as a meta-network. The question of whether the addition of auditory, olfactory, and other modalities to a visual meta-network would make a qualitative difference to our experience should be answered affirmatively. Multimodal merging enables the essential qualification that an input is "visual" as opposed to tactile or gustatory. If we were lacking all main modalities except vision, submodalities like "motion" and "color" would probably come to be experienced as main modalities, given the lack of a wider multisensory horizon, but they would be deprived of their multisensory anchors. A larger, multimodal meta-network⁴ will allow unimodal patterns to be embedded in a

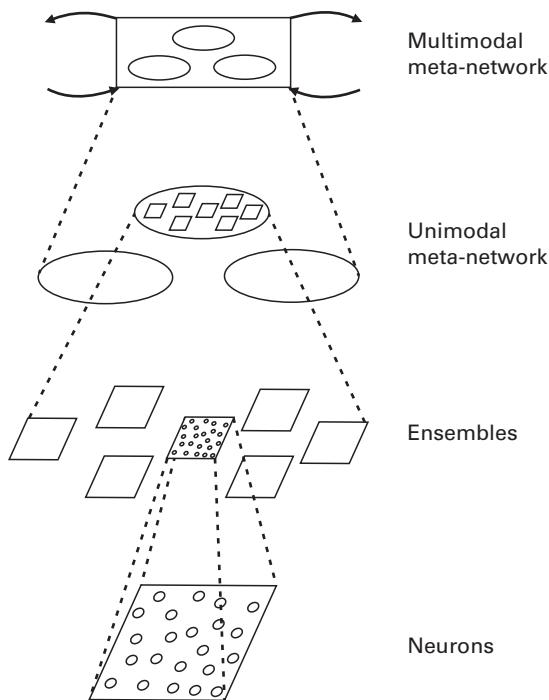


Figure 10.2

Functional delineation of organizational levels for understanding consciousness as a multimodal phenomenon. This delineation is different from a scale-based distinction of levels, because different levels are distinguished here based on *function*, i.e., what each level can accomplish in terms of computation and representation, culminating in perception at the highest level. Starting at the lowest level of single neurons, the next level is defined by collections of neurons (ensembles) with very similar firing behavior and engaged in the same computational function. Ensembles distributed across multiple brain areas that belong to a single sensory modality (e.g., vision) are hypothesized to communicate with each other, forming the higher organizational level of (unimodal) meta-networks. Integration amongst sensory modalities, but also between sensory and cognitive-mnemonic domains, is achieved at the highest level—that of a multimodal meta-network. Arrows toward and away from this meta-network indicate connections with other brain systems engaging in nonconscious processing such as motor systems. Connectivity between entities at lower levels is not shown here. Dashed lines between levels indicate relationships of correspondence, which are not causal in nature but do conform to the notion of “supervenience” (Kim, 1993, 1999; Van Gulick, 2001).

context defined by other modalities, and thus it is useful to maintain a distinction between these two types of aggregate level (unimodal vs. multimodal meta-networks). Also in this higher range of organization, there is no single level-to-level transition where inputs suddenly “turn orange.” Yet, the current assumption that phenomenal experience is attributed to the representational powers of this multilevel system has the interesting implication that its information processing is Janus-faced, as it contains both algorithmic and nonalgorithmic (phenomenal) aspects, which are by no means mutually exclusive. This view will be buttressed in the next section.

By itself, the delineation of four levels does not exhaustively address how systems for conscious versus unconscious representation differ in their organization. This question, also addressed in chapter 6, needs to be followed up in the current multilevel framework. For instance, in my own research group we have studied collaboration between hippocampal and ventral striatal ensembles (forming a meta-network in the current terminology) that are thought to function in the formation of place-outcome representations (Lansink et al., 2009; Ito et al., 2008). These representations are unlikely to contribute directly to conscious experience (cf. Pessiglione et al., 2008). One of the features likely distinguishing between nonconscious and conscious representational systems is the lack of intensive, bidirectional information trafficking. This trafficking is thought to occur between corticothalamic systems, enabling a constant updating of feature relationships in the multisensory domain and a concomitant “echoing” between sensory and semantic modules for cognitive interpretation. This is not more than a best guess, and a new field of computational and cognitive neuroscience is waiting to be mined here.

10.3 Symbolic Representations

Much of the controversy on the relationship between neural computation and representation *versus* perceptual interpretation is centered on the question whether neurons (or their aggregates) form *symbols*, to be deciphered or interpreted by their counterparts elsewhere in the brain. A symbol is defined as an informational element that represents, or refers to, something else than its own perceptual appearance. In Pythagoras’s law ($x^2 + y^2 = z^2$), x does not refer to a literal crossing of two line elements but to the length of one of the sides of a triangle orthogonal to another one. This is the problem we face when thinking about neural aggregates: they are taken to represent *something*, but from their own activity pattern alone it is not evident *what* this is. If the pattern is transported to another, receptive brain area, interacting with other inputs, it also does not become clear what is being represented, even when a meaningful behavioral output is emitted.

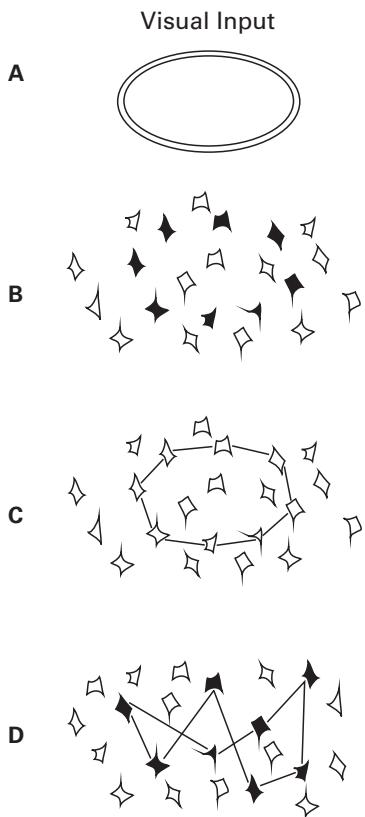
Departing from artificial neural network models, Smolensky (1987, 1988; see also Hofstadter, 1985) argued that neural networks such as multilayer feedforward nets operate at a subsymbolic and subconceptual level, but that they provide the lower-level, computational elements for higher-level, symbolic architectures. He conceived of mental

entities as higher-level structures operating by hard, symbolic rules, which are nonetheless isomorphic with lower-level, massively parallel computations that satisfy soft numerical constraints (such as used in neural network models). In Smolensky's (1988) view, consciousness would consist of the active maintenance of retrieved symbolic code, which results in a persistence of rule representation over time. However, similar to our discussion of the symbol-grounding problem (see chapter 5), it remains unclear in this account how symbols acquire meaning and, thus, how representation with phenomenal content would arise.

As soon as we choose to use "symbols" as a conceptual tool, we are bound to explain how they can be deciphered or interpreted (otherwise they should not be called symbols). To illustrate the situation, we suppose we have an ensemble code for some visual input, such as a ring or letter "O." Ignoring known feature-detector schemes in area V1 (e.g., simple, complex, and hypercomplex cells) and higher areas for the moment, we may deploy an ensemble in a given brain area that symbolizes the input isomorphically (see figure 10.3A, B). Different segments of a circle are represented by different neurons, copying the topographic relationships in the visual input literally. A ring-like structure can be obtained by having the ensemble's member neurons be more active than nonparticipating neurons (see figure 10.3B) or by enhancing the connectivity between the members (see figure 10.3C). Rapidly adjustable synaptic connectivity may be used to augment their functional connectivity (Von der Malsburg, 1995), possibly by way of nonzero phase relationships. So far, we merely "copied" the external O-shape onto an isomorphic representation in the brain. This representation is explicit to an external observer examining brain activity and contains information on the O-shape, but the local ensemble and its remote partners have no explicit knowledge of what it is representing.

Next, we make the ring representation less explicit while preserving the essential information: the same neurons stay active as they were before, but their positions in neuroanatomical space are scrambled (see figure 10.3D). Preserving the annular topology by way of functional connectivity and/or firing-rate elevations, the O-shape can still be recognized. Anatomical positions of ensemble members are essentially arbitrary: the rest of the brain cannot "see" these positions anyhow. Thus, the "neural symbol" for the O-shape is maintained, albeit invisible to an external observer unaware of member connections. In physiological practice, we will be dealing with tens of thousands of activated cells sensitive to many aspects of the ring as it is processed via various topographic maps. This parallel-distributed nature notwithstanding, the activation patterns still contain the same information as when the system was working in purely isomorphic mode. We can still cling to our beloved "neural symbol."

However, who or what could *decode* the symbol into something phenomenally meaningful? When, or by which mechanism, do we start seeing the "O"? The very concept of "symbol" presupposes it *can* be decoded or read out in some way—traditionally in a subsequent step of an information-processing procedure. Again, we have to get rid of

**Figure 10.3**

Neural symbols with explicit or implicit isomorphism. (A) External ring-like object, visually presented to a neural system (through various intermediate steps not shown here) such as in (B–D). (B) A neural symbol with explicit (directly recognizable) isomorphism, representing the visual input, can be produced by an annular arrangement of neurons with enhanced firing rates (black) relative to surrounding neurons. (C) Another mechanism for forming an isomorphic neural symbol is to indicate relationships between object-coding neurons by way of enhanced functional connectivity (e.g., phase relationships in firing). (D) Even if the neurons from (C) are scrambled throughout anatomical space, they maintain their (now implicit) isomorphism as long as their annular topology is preserved, in this case by functional connectivity. In general, isomorphism is defined as a preservation of structure: when one group of items is mapped onto a second group of items, the structure of relationships in the first group is preserved after transformation to the second group.

the idea that the symbols are out there in the brain waiting to be “read out” or translated into something meaningful by other neural structures operating *at the same organizational level*. Instead, I submit the only meaningful way to hold on to the notion of neural symbols is to have their interpretation occur at a *higher level* of systems organization.

10.4 The Wall Street Banker’s Argument

The visually presented O-shape is next rendered blue, thereby activating subsets of neurons in V4/V4 α and higher areas (such as the anterior–ventral portion of IT in macaques; Tamura & Tanaka, 2001). It is difficult (or even impossible) to imagine that a V4 ensemble codes for “blue” with its action potentials, because those action potentials are not looking blue and none of their physicochemical properties suggest why to interpret *them*, specifically, as coding for blue. But somehow our brain apparatus generating and processing this firing pattern is able to interpret it as “blue.” How could this happen? At the neural level, we note that some receiving brain areas will respond to the V4 activity pattern *consistently* in a blue-specific way, even when the receiver neurons themselves are better tuned to another modality, for example, shape, or act on the motor apparatus. The V4 ensemble activity will influence not only primary receiving areas but also topologically more remote nodes, such as the frontal eye fields, posterior parietal cortex, IT, and other temporal lobe areas (e.g., the parahippocampal areas TF and TH; Felleman & Van Essen, 1991; Banno et al., 2011). As long as we stay put at the neural ensemble level as our mode of description, nothing unusual happens.

Now we go up to a higher level of description, to that of uni- and multimodal meta-networks. Let us suppose that the active V4 ensemble *pretends* to code for blue. At the neural level we saw that such coding, as a gap-bridging principle per se, makes no sense, but let us for a moment indulge in the idea that the patterns in V4 and its coactive member nodes are somehow able to *enact* the color blue when considered from a higher-level systems perspective. The crux of the argument holds that this enactment of blueness can be regarded as an “illusion” that is never discovered by the brain (considered as meta-network) as being illusory. The brain therefore takes it to be real. The meta-network cannot discover the trick because it has no independent *access* to the information underlying the enactment, specified as it is by the low-level action potential patterns. The pretension of the collective neural activity pattern that the color “blue” is really there can be maintained because (at the meta-network level) there is no capacity to discover that a trick has been played. In other words, given the partitioning of brain function into different functional levels (see figure 10.2), activity changes at one level are posited to have *automatic* consequences for representation at other levels, but how these consequences take effect is *opaque* from one level to the other.

We still remember all too well how Wall Street dashed forward toward the 2008 financial crisis, with bankers selling packages of mortgage-backed securities strongly in excess

of their actual credit quality. Buyers received, so to speak, a “closed envelope,” with the actual value of the contents highly nontransparent to them. This is how the name of the above argument got coined, but two key differences should not go unnoticed. A subprime mortgage-backed construction can be backtracked to its roots, whereas a meta-network in the brain does not have access to this information. It cannot sort out that the basis of its blue-color symbols are “just spikes.” Second, Wall Street transactions can be understood as occurring at the same level of organization (person-to-person or firm-to-firm), whereas in the current account trans-level interactions are central.

The argument has further implications if we view it not from the bottom up (ensembles pretending or enacting) but from the top down (a supra-assembley network trying to make sense of low-level events). Which *cues* would the meta-network have to interpret an ensemble code in a certain way, for example, as representing the color blue? Without an explicit or implicit “key” for decipherment, the most reasonable way to go about this is to propose the meta-network interpret an ensemble code *however it likes*, although within certain boundaries. This is basically how subjectivity may arise: the brain’s *initial* interpretation of low-level inputs will be *arbitrary*, but once the brain has settled for (or “chosen”) an interpretation, the interpretation of other inputs is constrained by this choice. These constraints will be explored below. Of note, the meta-network level is not here to be introduced as a new homunculus in disguise, smuggling perceptual interpretation through a back door. There is no single neural–perceptual transition in this account, as interpretive capacities are built up gradually as we move up to higher levels of organization.

If we accept the premise, in trying to find a unified mind–brain theory, that there *is* a gap to be bridged (Levine, 1983; Nagel, 1974) and combine it with the above argument, an interesting postulate can be formulated. That is, if we are to explain the Janus-faced nature of experience, reconciling its roots in spike trains with phenomenal perception, the use and effectiveness of symbolic representations in and by the brain *has* to be accepted (as no other effective descriptors remain at our disposal to bridge the gap).

10.5 How Arbitrary Are High-Level Interpretations?

Because the current account assumes that higher levels of network organization have no independent information on how to interpret low-level events, this interpretation must be essentially arbitrary. But how then are we to avoid that low-level patterns become randomly mapped to any type of percept? That there is no objective, preordained way for the brain to decode low-level patterns into percepts fits well with the notion of subjectivity, but this should not be taken to imply that our percepts of 400- versus 405-nm wavelengths of light should be wildly different, given the same color context. Let us reflect on our options and first consider the argument that there actually *is* a strict relationship between an external input, a neural activation pattern, and its meaning. This is guaranteed by the specificity of sensory receptors transducing external physical energy into specific spike outputs and

by the preservation of this specificity during upstream subcortical processing. Unfortunately, this stance prompts us to fall back on the labeled-lines hypothesis—the specificity of activation per se is not sufficient for attribution of meaning. From the viewpoint of the rest of the brain that deals with the input-induced activation pattern, it remains an arbitrary (although unique) stream of spikes.

What else do we have, besides an exquisite set of feature detectors uniquely sensitive to specific sensory inputs? Arbitrary as the interpretation of activity patterns essentially is, the network's commitment to an interpretation of an input is required only once during the animal's ontogeny. As an animal is born and self-organizes its multimodal inputs into situational representations by which it learns to make informed behavioral decisions, it is sufficient for the brain to "decide" only once, in a physically deterministic sense, which set of inputs is henceforth taken to represent "blue" and so on. After this initiation, changes of interpretations become highly constrained, first of all by the *temporal consistency* in neural-perceptual relationships, imposed by the brain's relatively stable wiring. Blue-sensitive V4 neurons will influence target areas in a way that is consistently different over time from how yellow-sensitive neurons affect them. Assuming a constant center-surround context and absence of neural plasticity on the short term, repeated presentations of 450-nm light will elicit sufficiently similar ensemble patterns in V4 and its connected areas to yield a high-level "blue" interpretation distinct from one pertaining to 675-nm light ("red").

A further constraint arises from similarity-based consistency *within* a single representational system. If we compare similar inputs from the same submodality (such as cobalt blue and turquoise), a system would fail in its internal consistency if the corresponding representations were more dissimilar than for contrasting colors (blue and orange; cf. Clark, 2000). However, if we consider two *segregated* representational systems (e.g., the brains of two distinct persons) and expose them both to a 450-nm light pulse, *ceteris paribus*, then we are first confronted with the subjective and private nature of their experiences of blue, which may strongly diverge based on different personal histories, genetics, and interpretive commitments during early development. Secondly, and in contrast to Locke's (1667) inverted spectrum argument (proposing that your experience of a rainbow as a chromatic lineup from red down to violet may be the reverse of my experience of the same rainbow, going from violet down to red), the private nature of these experiences does not imply that colors can be swapped between persons at will, not even for colors having a symmetrical relationship on Newton's classic color circle based on chromatic similarity (see figure 10.4).

Once a mapping has been settled on, the *relationships* between colors (or other attributes) will not be coincidental as our representational systems obey the neural and psychophysical laws that apply here (e.g., von Helmholtz, 1866; Hering, 1874; Land et al., 1983; Palmer, 1999; Solomon & Lennie, 2007). A philosophical implication of the current theory is that two identical brains should experience the same sensory input identically (contra Putnam, 1975a; see also chapter 11).

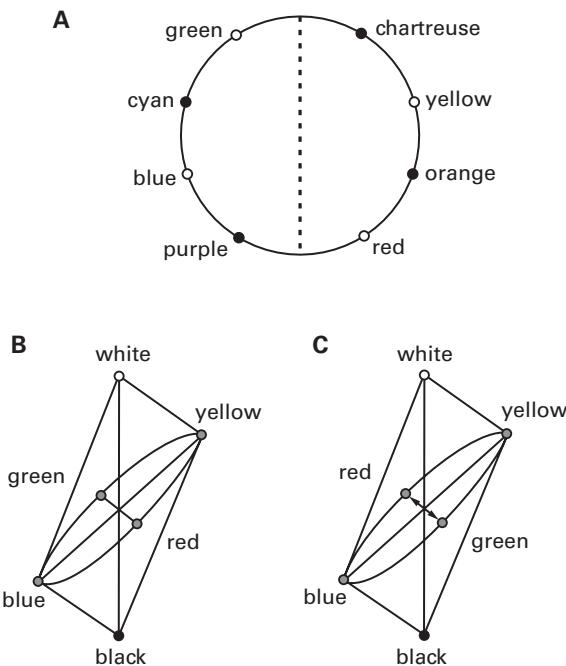


Figure 10.4

Limitations on inverted color perception. (A) Based on Newton's classic color circle, it seems conceivable that a person *A* would perceive certain chromatic inputs as, for example, purple, blue, and cyan whereas a person *B* would perceive the same inputs as red, orange, and yellow. This spectral inversion seems possible based on a vertical axis of symmetry in this circle (dashed line). This symmetry cannot be upheld, however, when taking into account the compositional relationships between colors. For instance, purple and orange can be perceived and analyzed as combinations of blue and red, and red and yellow, respectively, whereas red (as inverted from purple) and yellow (as inverted from cyan) are primary (unique) colors and thus cannot be decomposed. (B) A three-dimensional color space takes this compositionality into account. Its dimensions correspond to "lightness" (or brightness; vertical axis), "hue" (basic surface color; angular direction in the tilted horizontal plane away from the central white–black axis), and "saturation" (vividness; orthogonal distance away from the central axis, toward the green–blue–red–yellow ellipsoid). This psychophysically validated color space model imposes further constraints on the possible existence of perceptual color inversions that would not be behaviorally noticeable. For instance, a yellow–blue swap is discriminable because saturated yellow is judged as being lighter than saturated blue. (C) Some inversions, however, remain possible (e.g., red to green). Adapted from Palmer (1999).

In conclusion, for a well-behaved representational system, it does not matter whether a neural activation pattern represents an external world feature *truthfully* or not, as long as the system represents the feature *consistently* over time and in sensory space. As long as the representation remains in check with the input and its context of corepresented attributes, there is no rationale for the system to overhaul its interpretation—unless new evidence would become available falsifying it. Thinking of the neural representations our brain generates while contemplating Magritte's famous "*Ceci n'est pas une pipe*," I feel the theory could hardly be better illustrated than by this painting.

10.6 What Is the Nature of Relationships between Levels?

How can we best characterize the relationships between levels of representation? Do they interact with each other—and if so, could these interactions be of a causal nature? Our tradition of dualist Cartesian thinking prescribes that acts of the will, as outcomes of “mental” deliberation, exert effects on the body, and vice versa that stimuli in the bodily domain cause sensations in the mind. More recently, networks of nonlinearly coupled oscillators have gained popularity as a paradigmatic example of self-organization and mind–brain relationships (Thompson & Varela, 2001; Bressler & Kelso, 2001; Haken, 2006; Freeman, 2007). Neural groups constituting such networks engage in positive and negative feedback interactions, which give rise to emergent processes. These are understood as whole-system processes and cannot be (readily) explained from the properties of the system’s individual elements. Emergent processes are postulated to go in two directions (Thompson & Varela, 2001). An “upward causation” takes place when local system elements determine the global outcome, such as a massively synchronized state of gamma oscillations in a network. In “downward causation,” whole-system characteristics constrain local behavior of neural elements. For instance, Buzsaki (2006) describes how a collective gamma rhythm emerges by way of all individual contributions of single neurons in a network (upward causation); no single neuron plays a critical role in this. Once the network is in a synchronized oscillatory state, inhibition from multiple GABAergic interneurons constrains the timing of firing of individual neurons (downward causation). Inspired by this paradigm and the excitement about potential neural-synchrony correlates of perception in the 1990s (see chapter 7; Rodriguez et al., 1999), Thompson and Varela (2001) proposed that brain systems show reciprocal relationships between neural events and conscious activity.

The approach is inspiring and intriguing but nonetheless subject to critique. The examples provided to support the case concern systems phenomena that can be described at a “lower” and “higher” level (neuronal *vs.* network phenomena), but these levels refer to *scales* of organization (e.g., microscopic *vs.* mesoscopic) rather than conceptually different levels such as “neural” *versus* “perceptual.” The two types of causation are both manifested in the physical domain, and it remains unclear how conscious activity would arise from them. One may install as many nonlinearly coupled oscillators as one likes, but the transition from global network behavior (such as massive oscillation) to, for example, scene construction remains unexplained in this account. The trap of panpsychism is always lurking in the background, if only when considering other coupled-oscillators in mechanical (e.g., Huygens’s pendulum), electronic, and optical (e.g., laser) systems. Although this critique might backfire on the current account as well, I argue that this is not, or at most partially, the case considering our efforts to concretely delineate a *functional* division of multiple levels of representation (see figure 10.2) and further wheelwork to make a conscious mind’s clockwork tick (see chapter 9).

Theoretically, there is an important distinction to make between the two accounts, because a framework resting on reciprocal and *causal* interactions between neural and conscious domains is arguably a case of (covert) dualism (admitting consciousness as a “causally efficacious participant”; Thompson & Varela, 2001, p. 425). In contrast, as far as causality is concerned, the current framework is more in line with monism (Spinoza, 1677), but at the same time avoids panpsychic connotations. I would submit there are neither physical nor causal interactions between various levels (see figure 10.2) because the level-to-level relationships are defined functionally, by their representational structure and computational scope. A significant change at the neuronal level *automatically* corresponds to a change at higher levels, not because it exerts a causal effect on those levels but because all four levels reflect the same mind–brain state. This agrees with the philosophical notion that a conscious representation supervenes⁵ on the lower, neural level (Kim, 1993, 1999; Van Gulick, 2001) but goes beyond the prior art in specifying intermediate levels.

An analogy takes us back to the Republic of Florence in 1494. The power of the liberal Medici family was declining, and the zealous Dominican friar Girolamo Savonarola took control of the city-state, ordering destruction of a cultural legacy consisting of works of art, books, and cosmetics. Can we say that the physical destruction was the *cause* of the cultural loss? The physical and cultural changes refer to the same event, and the cultural destruction happens simultaneously with the physical destruction, without causal interaction. Similarly, the loss of color vision in a subject’s brain can be said to be caused by an embolism in the artery supplying blood to area V4–V4o. However, loss of neural function in this area automatically, without any time delay or intermediate causal step, corresponds to the loss of color vision. Perceptual processes defined at the unimodal and multimodal meta-network levels and neural phenomena are flip sides of the same coin.

Could we then maintain that consciousness—manifested at the highest levels of representation—has a separate *function* over and above neural phenomena? Earlier on, I argued its function is to provide a multimodal situation sketch of what goes on in an organism’s immediate environment and body (see chapter 6). At present, it may seem that consciousness is reduced to an epiphenomenon and the real hard work is done by all those little neurons. However, if we accept that the lower, intermediate and higher levels exert no causal effect on each other—that the type of relationship they have is not of the form “A influences B”—then the distinction between epiphenomenon and “real work” vanishes. Multimodal representation is a process realized simultaneously at all four levels. What distinguishes the various levels can be best described as representational or experiential complexity, and it is an open question for future research to flesh out this description using empirical data. At the level of multimodal meta-networks the system has the full representational power to smell and see a rose. At the unimodal meta-network level experiential complexity is lower, leaving only one modality to provide a dimensionally reduced sketch of one’s surroundings and body state. At the ensemble level, the distinction between submodalities is lost, leaving us with within-area activity bumps and attractors. Further

dimensional reduction takes place when descending to single neurons. The *function* of conscious representation is exerted at all levels, and no level is more epiphenomenal or superfluous than another—although the experience can only be “felt” in full force when the highest level functions properly.

We could consider the four different levels *equivalent*, were it not for this term’s implying they would be of “equal value.” I will refrain from using this term because the quality of rose smell cannot be measured along the same scale as used for quantifying spike patterns. Low-level neural phenomena are defined in the domain of physics, which deals by its very nature not with qualitative phenomena, such as subjectively perceived color, but with wavelengths, photoreceptor transduction of light into electrical signals, and interactions such as center–surround inhibition. If we were to keep looking for an exact *equivalence* between neural processes and experiences, we would be bound to fixate on the gap over and over again: purely on its own, the language of physics does not allow one to *demonstrate* how qualitative phenomena arise from brain activity patterns. Therefore I propose to use the term “correspondence” to indicate the nature of the immediate, noncausal relationships between levels (see figure 10.2). These relationships are nonetheless nonsymmetrical because the higher-level functioning is supervenient, or dependent, on the lower levels, but low-level elements such as neurons can in principle function on their own (e.g., in a culture dish) without being dependent on high-level properties.

The distinction of four levels may give us a conceptual handle for further investigations but does not imply that we know how to exactly describe conscious experience at a meta-network or other level. A question for future research is to uncover how phenomena and relationships at a single level and between levels can be characterized in terms of rules: what are regularities in correspondences; what is the “syntax”? With a sufficiently large physiological toolbox, preferably including probes to measure large-scale synaptic connectivity, we may physically examine and describe a meta-network state, but to *experience* this state, one would have to *be* that same meta-network. This would be possible only if all measurements could be done in our own individual brains. Given moral and practical objections, we will have to satisfy ourselves with behavioral reports on conscious experience in nonhuman animals having a brain sufficiently similar to ours.

10.7 Impenetrability between Levels: The Limits of Imagination

Suppose you would, by some miraculous novel technology, be enabled to map the activity patterns of myriads of neurons in your brain and see how their activity goes up or down together with your percepts. All of these activities would be plotted in a giant, three-dimensional diagram. To capture neural dynamics, we would compose a movie showing how neurons respond while you are watching objects and listening to music. You might stare at this jaw-dropping motion picture of your own brain for a while, but as time passes, you are beginning to get bored following the ephemeral patterns of dots lighting up and

extinguishing, and you murmur, “All well and good that you figured it out like this, but *that* is not what I am experiencing!” With abysmal proportions the explanatory gap appears to open up in front of us once again.

This would be a natural reaction to the situation. However, there is a powerful counter-argument: our imaginative capacities are too limited to see the multilevel correspondence between experience and neural activity. Imagination will take us only as far as perception itself will go. This means that the levels situated below the multimodal meta-network level at which our daily perceptual stream unfolds are not accessible for inspection. In his psychosemantic framework Fodor (1983; chapter 5) coined the term “cognitive impenetrability” to denote that a given computational module has no access to modules operating in another functional domain. Here, “impenetrability” is used differently, because Fodor proposed modules to work at the same level of organization (for instance, modules for vision vs. those for reasoning and abstract knowledge). In contrast, the current account places impenetrability at the level of phenomenal experience, leaving no room for a keyhole through which we can secretly watch what is going on at lower levels. We cannot look under the hood of imagination to see what is going on in the engine. Returning to the movie displaying our own neurons lighting up during streams of experience, it makes complete sense that we cannot directly relate our subjective experience to it. This is not a hand-waving statement: it points to a fundamental epistemological restriction in our knowledge system and thinking. Whatever explanation we come up with to bridge the gap, we will never be able to conjure up a color experience from *that* (a neural configuration of activity).

Revisiting the Wall Street banker’s argument, another, potentially better metaphor may be introduced to drive the point home. We are presuming a robot has been built that is specialized in identifying the species an animal belongs to based on its behavior, and it does so using low-resolution video images. An ant-like minirobot is crawling in front of the robot’s eye cameras, precisely mimicking a real ant’s behavior.⁶ Naturally, the robot will report it detected an ant, unable to grasp it was exposed to a biomimetic artifact. Because the minirobot *behaves* like an ant, it *becomes* an ant to the robot, which has no way of detecting that the minirobot’s legs, head, and antennae are made of nuts and bolts. Similarly, a meta-network has no way of seeing what processing goes on at a low level of organization. Here, ant-like behavior is comparable to the perception of color and other qualities. Also the brain cannot identify a minirobot (as a conspiracy of ensemble activity patterns) as something other than an ant, because it does not have an independent reference specifying what the object might be other than an ant. This *lack of external reference* has significant consequences to be explored in the next chapter.

David Chalmers (1995) has criticized previous attempts by neuroscientists to identify neural correlates of consciousness. For instance, when coherent gamma-oscillating firing patterns of visual cortical neurons in response to coherently moving line segments (Gray & Singer, 1989; see chapter 7) were proposed as a potential neural correlate of consciousness

because of its presumed role in binding (Crick & Koch, 1990), Chalmers (1995) made the following riposte:

Why do the oscillations give rise to experience? The only basis for an explanatory connection is the role they play in binding and storage, but the question of why binding and storage should themselves be accompanied by experience is never addressed. If we do not know why binding and storage should give rise to experience, telling a story about the oscillations cannot help us. (pp. 204–205)

Earlier on, we encountered empirical arguments against gamma synchronicity as a neural correlate of consciousness. However, the point here is that Chalmers's argument could be applied to reject *any* potential neural correlate *C* suggested by neural recordings because *C* is described in physicochemical terms and thereby is unfit *per definition* to provide a conclusive explanation. The counterargument was already laid out: let us not overrate the power of our imagination, incapable as it is in jumping across multiple representational levels simultaneously. No matter which neural mechanism we come up with, and no matter how complex or powerful it is, it is impossible to show *directly* how phenomenal experience corresponds to patterns of neural activity. What we *can* do is to situate the suspected neural correlate at the right level of processing and flesh out its relationships with phenomena defined at higher levels by doing a great deal of laborious research.

In summary, I have argued in this chapter how Marr's multilevel notion of mind–brain organization can be modified to accommodate the functional and representational demands applying to conscious brain systems proposed in earlier chapters. Concretely, the modified account ranges from the single-neuron level to functional ensembles (for which empirical evidence is emerging) and, hence, to unimodal meta-networks that are integrated into the higher level of multimodal meta-networks. No saltatory transitions in consciousness or neural-to-mental activity exist between levels, as meaning and experience are built up gradually when moving up to higher levels. The relationships between levels can be best described as noncausal, with higher-level phenomena corresponding to, or supervening on, lower-level phenomena. Crucially, higher-level representational entities have no independent access to lower-level processing: what is experienced at a higher level as having meaningful content cannot be directly unmasked by the same subject as being merely “neural” at a lower level. A fruitful approach is to view meaning, situated at a high level, as arising from symbols, implemented by ensembles, in a way that allows ontogenetically developing brain systems to make arbitrary representational choices within constraints defined by temporal consistency, the structure of sensory space, and other psychophysical aspects.

11

Philosophical and Future Perspectives

A mother told me she was suddenly aware, as my electrode touched the cortex, of being in her kitchen listening to the voice of her little boy who was playing outside in the yard. She was aware of the neighborhood noises, such as passing motor cars, that might mean danger to him.

—W. Penfield (1975, p. 21)

11.1 Comparison to Philosophical Schools of Thought—Dualism

In this final chapter we will reflect on some of the philosophical implications of the theory presented in previous chapters, as well as its consequences for future research. Without any claim toward completeness, I will attempt to position the current account relative to other schools of thought, such as dualism, materialism (and the cognate positions of functionalism and externalism), and representationalism. These views can be demarcated using some classic philosophical benchmark problems like “Mary the Color Scientist” (Jackson, 1982) and the case of zombies. As to matters of high societal relevance, we have space to face only a few of the many questions that have come our way. Can animals generate consciousness—and if so, what properties should nervous systems in “lower” species minimally have to do so? Is there a sharp, identifiable threshold in nervous organization, or do the nonconscious and conscious processing of environmental inputs merge gradually into one another? In a similar vein, our daily lives have become increasingly crowded with complex electronic and informational devices such as computers, smartphones, robots, and global Internet phenomena, and so we are confronted with the question whether these human-made inventions possess sufficient complexity to generate conscious representations. What would be a critical litmus test to validate these ideas?

Finally, a great number of scientific and ethical issues call for attention when brain functioning massively goes awry. Injury and stroke can bring brains into a comatose or vegetative state, raising the question of under which conditions consciousness may be preserved or could otherwise be restored.

In the motley procession of philosophical theories dominating the contemporary landscape of mind–brain debates, only classical dualism is notoriously absent. By now the

Cartesian concept that the mind and body are two different substances of a very different, incomparable nature—which nonetheless mysteriously communicate with and influence each other—is no longer tenable. Although classic dualism is still highly influential in our society and daily conversation today, this position has been unmasked as deeply problematic in many excellent sources on this topic (e.g., Block, 1980a; Searle, 1992, 2004; Rose, 2006; Dennett, 2005; Alter & Howell, 2012).

A more modern version is property dualism, which is different from the classic variant in that it does not conceive of mind and body as two different substances but maintains that mind–brain systems have two fundamentally different properties—physical and mental properties. On the one hand, they deal with light waves of 540 nm and, on the other, sustain mental properties of the greenness of a perceived apple or the belief that a green apple is good for your health. This view is not committed to the existence of a separate, nonphysical substance. Nevertheless, property dualism faces some of the same problems as substance dualism (e.g., Searle, 2004; Rose, 2006). If we abandon the notion of two separate substances but maintain one substance with two radically different kinds of property, how is it that these properties, or the objects that these properties are attributed to, relate and interact? So far, property dualism has not explained how or why a 540-nm light wave would give rise to a phenomenal sense of “green,” and the credibility of the position depends very much on the availability of a heuristic to find such an explanation. However, this heuristic, or road map toward a solution, has been lacking.

The problem is poignantly illustrated by asking how a mental, nonphysical property could exert a causal effect within the physical domain of object properties—how perceiving the color of an apple results in grabbing this object physically, using motor neurons and muscle fibers. We can fully describe the entire sequence from photoreceptor activation in the retina to muscle activations and hand movement in one uninterrupted physical loop, without needing to invoke any mental property interfering with this sequence. Replacing the term “substance” by “property” to denote a different type of dualism is not helping us forward. By consequence, mental properties would be reduced to epiphenomena as they have no demonstrable role in organisms interacting with their environments.

11.2 The Zombie Argument

The zombie argument has acted both as a litmus test and as fission material in dividing philosophical opinions over the past few decades. Its recurring prominence also draws attention to more contemporary versions of dualism, monism, and related positions (Nagel, 1980; Chalmers, 1995, 1996; Searle, 2004). Is it conceivable that there would be an agent that looks and behaves exactly like you do but is devoid of any conscious life? If we can imagine such an agent, so this argument against materialism goes, then mental states must be kept apart from physical brain states. Importantly, proponents of the argument claim it

is sufficient that a zombie be *conceivable*, or *logically possible*, to conclude that mind and brain are fundamentally different entities (e.g., Chalmers, 1996).

Suppose we undertake this enterprise by a tremendous feat of neuromorphic engineering. We are calling Harry's surgeon (see chapter 3) back to the operating room and asking him to replace all the neurons in Harry's head with electronic chips that will *exactly* imitate these very cells—in their synaptic connections, plasticity, dendritic potentials, and spikes—as precisely timed and shaped as their originals. Is it logically possible, then, to carry out such a full replacement but without endowing the system with conscious capacities? It certainly seems *possible* to imagine or conceive this. We could even make our experimental lives easier by not designing all kinds of neuromorphic chips but by replacing brain tissue molecule for molecule. If we would do that, would it be logical to infer that we are just reconstructing a molecular machine without any phenomenal experience? When comparing himself to a zombie, Chalmers (see <http://consc.net/responses.html#balog>) states the argument against materialism as follows:

The justification for my belief that I am conscious lies not just in my cognitive mechanisms but also in my direct evidence; the zombie lacks that evidence, so his mistake does not threaten the grounds for our beliefs.

This argument highlights where some of the schisms and misunderstandings in the field are. First, nature is carrying out the “molecular replacement” experiment on our brains all the time. All of our ion channels, membrane lipids, and potassium ions are continuously subjected to a biochemical turnover, just like in the rest of our body—a process taking roughly 1–5 years to complete. In that sense, the thought experiment is very real and we can decidedly confirm that we do not have to give up our precious conscious experience while the replacement is ongoing. However, the crucial question was this: is a zombie *logically possible*? Is the “mere intelligibility of the notion” (Chalmers, 1996, p. 96) a sufficient argument for the irreducibility of consciousness?

The answer is that the question is ill posed and does not properly apply the term “logical.” Logic is the science that examines the structure and validity of formal propositions, arguments, and inferences—not by examining their contents empirically but by checking whether they satisfy strict rules, following mathematical formalisms. This approach is not at stake in the zombie argument. If we regress to asking whether zombies are *conceivable*, we are not studying a logical problem but instead come to rely on our imagination. We can observe all kinds of machines with interesting behaviors, such as the vacuum-cleaning robot Roomba or Sony's dog-like robot AIBO, with little reason to suppose a “mind” in those machines—but it is not feasible to imagine the *absence* of some hidden property or entity—a mind or “animal spirit”—in such an agent. The absence of some feeling or other phenomenal experience cannot be the subject of our imaginative efforts. It is not the business of our imagination to do this. We can imagine all kinds of things, from pink elephants to flying saucers, but our imaginative capacities are limited to the reconstruction of sensory

experiences, or to novel constructions based on previous experiences. Coming back to Chalmers's argument, I do not agree that we can say that the zombie would lack the direct evidence that I would have myself as a conscious being. We do not directly know—and cannot imagine—what the zombie experiences himself (as the zombie should be able to speak up, via its sensorimotor mechanisms, and assert that he does have "direct evidence"). If the zombie is an exact molecular copy of yourself, all the neural means to gather direct, experiential evidence would be the same as in a conscious "I," so the grounds to deny this evidence to the zombie are lacking.

However, if we revisit a neural network and conceive of neurons connecting to each other and emitting flashes of activity—would it not be reasonable to say all that activity is possible without conceiving any mental life in the network? Again, "conceiving" refers to an act of imagination. What would you think of when doing this—other than retrieving sensory (including auditory–linguistic) experiences from the past, or recomposing them? When you recall such experiences, like catching a blue ball, does it make sense to say that you situate them "in that network"? Recalling the limits of our imagination, we cannot look under its hood to check directly how neurons are busy coding a blue ball. We equally fail in "bridging the gap" between mental and neural levels because the concept of having a spatial divide between these levels is erroneous unless applied weakly as a metaphor. We do not know exactly when neural relationships give rise to conscious or unconscious representations, but arguments relying on imaginableness or conceivability will not help us further.

In sum, it seems theoretically possible and conceivable that zombies exist. However, they are not realizable. The situation is somewhat comparable to a hollow planet, which is certainly conceivable but not realizable because of the law of gravitation. If your brain were to be replaced molecule by molecule, your neurons, networks, and their larger aggregates would continue to function just like they did before, and the same is true when replacing neurons with chips exactly copying their functionalities. Conscious representations will be inevitably generated by these configurations of activity because their representational functions are also preserved. Nonetheless, these representations are not epiphenomenal, as they do not stand in a relationship with the neural activity configuration that is considered causal in a conventional sense.

Generating the exact neural patterns appropriate for conscious representation, but without actually having those representations, requires an (imaginary) *deus ex machina* that would come to the rescue of a dualism as advocated by Chalmers. The previous chapters have only given a scant summary of the extensive neuroscientific evidence that there is no such additional factor—a mysterious "vitalistic" force existing next to or beyond known neural principles. There is also no theoretical necessity to suppose its existence.

11.3 Materialism

Given the problems of dualism, and the successes of physics in explaining a wealth of natural phenomena over the past four centuries, the advent of materialism (or physicalism) is not surprising. Materialism acknowledges only physical entities and their interactions as the ultimate ground of reality and thus tries to explain mental states in physical terms, grounded in the laws of gravity, electromagnetism, relativity, quantum theory, and so on. Most variants of this program are reductionistic in that they attempt to decompose mental phenomena into underlying neural components. For instance, an act of will is reduced to the dynamic evolution of neural activity patterns in the frontal cortices and basal ganglia.

A strong variant of this school is eliminative materialism, which attempts to do away with mental phenomena and psychological descriptors completely (Feyerabend, 1963; Rorty, 1971; Churchland, 1984, 1995). It argues that concepts like mind, mental states, beliefs, and consciousness are inherited from an outdated folk psychology that has no solid scientific foundation. A distinction between “mental” and “physical” states is as nonsensical as the medieval notion of the sublunary and the superlunary. Because there is no independent scientific justification for “mental” phenomena as distinct from physics, we should stick to the demonstrably successful program of physics, abolishing psychological jargon.

A second important variant of physicalism is identity theory, also known as type-type physicalism (Boring, 1933; Place, 1956; Smart, 1959; the variant of token-token identity theory is not treated here). Different from earlier physicalists, who held that dualism was untenable because of logical inconsistencies, identity theorists argue that the reducibility of mind to brain is a matter of fact, evidenced by empirical findings demonstrating that mental processes strongly correlate with brain processes. This theory groups conscious experiences into types (or classes, such as “seeing red”), which have exact correlates with certain types of physiological events in the nervous system. In a classic example, the category of “pain perception” would correlate exactly with action potential firing of C-fibers, which are a particular type of axon in the somatic sensory pathway conveying information from peripheral body parts to the spinal cord’s dorsal horn.¹

The current theory has a take on the problem that differs from both eliminative materialism and identity theory. It views “mental” phenomena as situated at a high representational level, occurring exactly simultaneously, without any time delay, as events situated at lower representational levels (see figure 10.2). A “mental” and “neural” phenomenon refer to one and the same event but described or represented in different ways. The relationships between levels are characterized as *corresponding* to each other, avoiding the notion of causality. By consequence, any level is not more or less of an epiphenomenon than the other level. There is no need to explain how “mental” and “physical” properties or entities interact because the whole idea of inter-acting, as an act of elements in one system affecting each other, does not apply. Explaining how this level-to-level correspondence works

out exactly is still in its infancy, but nevertheless this framework is argued to offer a better heuristic for conceptual progress.

The situation is slightly comparable to the notion of liquid water freezing to ice when cooled down. The macroscopic quality of solidity arises simultaneously as the microscopic arrangement of molecules into a rigid lattice structure, and there is no causal interaction between these macro- and microscopic levels of organization; we just describe (or visualize, represent) the same phenomenon in different ways. However, does this comparison not force us to admit that mental phenomena are, in fact, *reducible* to neural phenomena? The argument here is that a liquid-to-solid state transition of water is only partially analogous to mind–brain phenomena. Much depends on what exactly we mean by “reducible.” All properties of a freezing volume of water can be theoretically explained by molecular and thermodynamic principles, so in a very real sense the macroscopic level of explanation becomes superfluous, as macroscopic events can be reduced to molecular phenomena (cf. Kim, 1999). However, for consciousness, a similar reduction does not apply. Many cognitive properties (e.g., memory, attention) can be explained, in principle, by neuronal mechanisms and supraneuronal organizational principles, but this does not mean we can do away with all “mental” vocabulary as being superfluous—we cannot avoid talking about qualitative phenomena and cannot reduce the taste of sauerkraut to bits and bytes. In this special case, explaining or describing is *not* equivalent to reducing (cf. Bechtel & Richardson, 1993; van Gulick, 2001).

Overall, the multilevel representational scheme presented here does acknowledge the reality of phenomenal experience as much as of physical events such as spike patterns. Eliminative materialism will not work, because the basic language of physics is not equipped to describe subjective experience appropriately. It is reasonable to invoke *some* psychological vocabulary—at least the vocabulary minimally needed to describe or explain the richness of mind–brain phenomena we believe are fundamentally shared between humans, insofar as it cannot be captured by the terminology of physics. Even stronger: the language of physics was designed to describe our world in measurable, quantifiable terms, and this approach has defined its success—but a price had to be paid in that references to some subjectively experienced qualities had to be excluded.

Coming back to identity theory, its explanation of mental phenomena as exact correlates of certain kinds of isolatable physiological events is rejected on similar grounds. Precisely the example of C-fibers and pain perception illustrates the deficiencies in theories relying on a labeled-lines assumption and seeking *simple*, isolated correlates of consciousness in the nervous system. Even if such a single, correlative phenomenon could be identified, it would by itself not *explain* why anything like a qualitative experience will ensue (an argument reminiscent of the panpsychic trap). Additional arguments against identity theory can be found in, for example, Kripke (1972), Chalmers (1996), and Searle (2004).

11.4 Functionalism

As a branch of materialism, classic functionalism casts mind–brain processes as causal intermediates between sensory inputs and behavioral outputs (see chapter 3). A state of “feeling cold” is not defined by an experience with phenomenal content but by its causal role, standing in between a gust of chilly wind and a shivering or the utterance “It’s freezing cold!” One of the factors that made functionalism so irresistibly attractive to many is the apparent match it offered between mental processes and computer algorithms. Also a computer’s computations act as a causal relay between input and output, and the analogy to the mind–brain is, at least, seductive. Here we can only briefly review the most relevant arguments on functionalism and see where the current theory stands in this debate.

In chapter 4 we revisited Searle’s (1980) Chinese room, arguing how the application of rules to convert input symbols into outputs does not give insight to the person inside the room as to what kind of information is being handled. Functionalists countered the argument by replying that one should not focus on the person inside the room alone but situate the “understanding” of Chinese at the systems level. This level includes not only the little man inside but also the room and the instruction toolbox. The well-taken riposte of Searle was that adding these peripheral systems does not make the difference, because they will not help in getting to understand the Chinese symbols when you are in that room (Searle, 2004). Syntax does not convey semantics.

Throughout the past decades, the lack of content, meaning, or phenomenal experience in input–output models has indeed proven to be the functionalist’s Achilles heel. If a mental state is purely regarded as a causal intermediate between a sensory input and an externally observable behavior, there appears to be no ground left for qualitative aspects of experience such as feeling pain or tasting chocolate. Functionalists have attempted to reduce the infamous “qualia” to discriminative states associated with belief states and emotional dispositions, but in chapter 3 we saw how, in clinical practice, conscious perception is inferred to persist when emotion-related brain structures are damaged. Emotion can be very well dissociated from perception or phenomenal experience, and so there must be something else about “discriminative states” that makes us feel and experience something. The same reasoning applies to beliefs, thoughts, and language.

The point is also made by the spectral inversion argument, which can be illustrated by two healthy persons looking at a grass lawn and stating how “green” it looks to them. The physical inputs to their eyes and their behavioral utterances are the same, yet their phenomenal experience can be different. What “seeing green” is for one person could be experienced as “seeing red” by the other (see figure 10.4). Phenomenal experience is ignored or reduced in functionalism, yet it is very much what makes up the essence of our conscious lives and needs an account that accepts the existence of its qualitative properties (cf. Shoemaker, 1990). Thus, the main objection against functionalism is that functional and physiological descriptions leave out something essential—something qualitatively

different from the spike trains running through one's brain—namely, a phenomenal experience that is as real as its physiological counterparts (Chalmers, 1996; Searle, 2004).

Having said this, functionalism also deserves credit from the viewpoint of neuroscience and computation. For instance, it alerted a broad audience to the idea of multiple realizability. In analogy to computers, functionalists reasoned that it should be possible in principle to have the brain's computational operations carried out by biomimetic computer chips, and I see no convincing argument to deny this proposal. As far as we know, all functions and operations of "wet" neurons can be implemented by classic electromagnetic devices (with the possible exception of quantum mechanical properties of ion channels; Zhong et al., 1998). Nevertheless, this notion must be kept separate from the main tenet of classic functionalism, namely, that mental or phenomenal states are reducible to causal intermediates in input–output chains of physical and behavioral events. Multiple realizability is perfectly compatible with the notion of phenomenal experiences as real, irreducible phenomena, which in the current framework are explained as corresponding to events unfolding at lower representational levels.

11.5 Externalism

A recent shoot on the physicalist tree is externalism. Advocated by O'Regan and Noë (2001; O'Regan et al., 2005; see also Ryle, 1949; MacKay, 1962, 1973), this view proposes that conscious experience results not from "internal" representations in the brain but from the interaction between an organism and its environment. Not unlike behaviorism and functionalism, externalism posits that the brain functions as an intermediate to couple sensory inputs to appropriate motor outputs but also emphasizes the environment as a codeterminant of conscious experience. This view is buttressed by detailed examples from psychophysics. The difference between specific experiences set in various sensory modalities is made up by the rules by which motor actions govern sensory changes. These rules are alternatively labeled "governing laws of sensorimotor contingency" (O'Regan & Noë, 2001, p. 939). When we rotate our eyes, the pattern of impinging light shifts across the retina in a way that is unique for the visual system, not found in other modalities. Seeing the color of an object in the visual field is equivalent to knowing which structure of sensory changes the color brings about. Thus, conscious experience is considered a way of acting on and interacting with the external environment, not an "internal" brain event that makes us see or hear, even if motor interactions with the environment were absent.

A merit of externalism is that it brings motor systems and psychophysics back into the spotlight when trying to explain the differentiation between sensory modalities. Also chapter 9 proposed a central role of the motor domain in calibrating and aligning signals from different modalities relative to each other. However, a problem of externalism is that motor-related signals are only one out of many sources by which modalities can be individuated. When motor behavior is persistently absent, such as in paralysis, modalities

continue to be distinct. The importance of sensorimotor contingencies might not have been fully recognized until now, but the knowledge of these contingencies, applied while we explore the world, is highly implicit. It is “knowledge” embodied in the biomechanical principles governing the physical execution of eye- and other sensor movements—in the hardwired connections between brain structures mediating largely reflexive sensorimotor conversions. Because of this hardwired embedding and automaticity, it is largely unrelated to conscious perception. Perception is maintained by neural activity patterns of *very particular kinds*, as I have argued, and disappears as soon as these patterns fall silent and neurons go back to a dormant or resting state. Many neural systems, including the cerebellum and basal ganglia, are crucial for setting sensorimotor contingencies but do not generate the right kind of patterned activity for conscious representation (see chapter 6).²

A further problem for externalism is that conscious experience can proceed in the absence of not only motor activity but also changes in sensory inputs. Stimulation of sensory cortical areas by intracranial electrodes or transcranial electromagnetic pulses can elicit percepts such as phosphenes (e.g., Cowey & Walsh, 2000; Antal et al., 2004). One may object that such stimulations will hit on pathways implementing a previously established sensorimotor contingency, but this is countered by the finding that these phosphenes are not usually coupled to established or spontaneous motor responses. Other major cases of experience without external input are imagery and dreaming. During dreaming, our bodies are largely paralyzed. Patients that are paralyzed by a spinal cord transection, who neither receive sensory signals from their genitalia nor exert control over them, report dreams with orgasmic sensations (Rechtschaffen & Siegel, 2000). O'Regan, Myin, and Noë (2005) rebutted that, in dreaming, imagery or other conscious feelings, “what counts in giving the particular “sensory” feel of sensation is not the actual sensory input itself, but the reliance on *implicit knowledge* that the sensory input possesses corporality and alerting capacity”(p. 377). In addition to the above critique, it should be noted the emphasis shifts away here from actual sensory input to the state of *expecting*, implicitly, what will happen if, for instance, one's eyes were to move, but do not actually move. The difficulty is that the content of experience is determined not, or at best partially, by such conditionalities or potentialities, but to a major extent by actual sensory input (“I'm surprised to see you wearing a blue scarf”).³

Even patients that have been locked-in or paralyzed for many years do not report that their conscious experience deteriorates or becomes blurry (see chapter 3; Bauby, 1998; Laureys et al., 2005; Lulé et al., 2009). Limb amputations are often followed by phantom sensations for much of a patient's life (Ramachandran, 2005), obviously without the option of exerting the relevant sensorimotor contingencies. Instead, these sensations are plausibly explained by the fact that their synaptic matrix for limb representation in the somatosensory cortex, albeit plastic, remains present and turns active from time to time, resulting in configurations at the ensemble and meta-network levels corresponding to percepts. The most dramatic example refuting externalism comes from a woman who reported sensations

in her limbs without any possibility to have them physically interact with the environment (because she was born without limbs; see chapter 3; Brugger et al., 2000).⁴

11.6 Mary the Color Scientist

In contrast to the zombie argument, Frank Jackson's famous knowledge argument (Jackson, 1982) does present a case that can be empirically tested in principle and is highly informative when evaluating the variants of materialism reviewed above. Mary is a neuroscientist who knows everything there is to know about the physics of light and the neurophysiology of photoreceptors in the retina, the optic tract, thalamus, and visual cortices, and all organizational principles relevant to understanding how neurons and brains process color information. But—cruel or blunted as her educators might be—she was raised entirely in a black, white, and gray environment, and she never saw the blueness of the sky or the redness of roses. Jackson argues something is missing in Mary's knowledge, as she cannot know what the color "blue" actually means, what it looks like. Once she steps outside her gray world, she will see the color of the sky and thereby experience something new. She will gain knowledge. On the contrary, a materialist account would imply that a complete description of neurons, spikes, and so forth, would suffice to capture everything there is to know about colors. However, this entails the paradox that Mary would not learn anything new when seeing the blue sky for the first time. Hence, this account cannot be complete.

I find this argument quite compelling, as it is difficult to maintain that Mary would not have some kind of novel experience when she steps out of her colorless housing. Let us probe whether the current account is challenged by the argument. It poses multiple levels of mind–brain representation, beginning with the nitty-gritty of neurons, going up via single-area ensembles, and ending at the apex of multimodal meta-network function and corresponding perception. These distinct levels require different kinds of description. In this framework, only the highest level of representation in Mary's brain corresponds to her full-blown perceptual capacities (see figure 10.2). The subject Mary is actually a multimodal meta-network herself (apart from the rest of her body), and this network functions such that its mechanisms can only self-scrutinize its global level of experience, not the microscopic operation of its neural nuts and bolts (level-to-level impenetrability). Via the same high level of experience, Mary has gained sensory inputs coming from textbooks (etc.) pertaining to cells, wavelengths, and neural circuits. Thanks to her further general knowledge and capacity for abstraction, she has been able to form concepts about the lower levels of neural color processing. The observation that eventually our knowledge of physics also derives from phenomenological experience (plus a great deal of math and abstraction) is not insignificant.

When Mary thus steps outside her gray world, she will indeed have a novel experience. This reaffirms the point that accounts purely based on quantitative, physicalist, or functionalist descriptions cannot be complete (see chapter 10). However, this should not drive us to

embrace dualism again. Before leaving her enclosure, Mary faces a problem of experiential and epistemological access. Through her laborious study of physics and neurobiology, she has been forming a knowledge structure for grasping the low-level side of brain functioning but has not gained access to the real thing: self-experienced color. Stepping in the outside world suddenly opens her door to high-level representation of color. The inability to make or see a direct connection between this high-level phenomenology and the lowly neurons is no excuse to readopt dualism. Rather, the current account can be best circumscribed as a nonreductive, multilevel, neurobiological form of representationalism (but see below for other representationalist accounts). Mind and “brain-in-the-right-state” are the same entity, but our routes of access or descriptions can be different and seemingly divergent.

The validity of Jackson’s argument is supported by empirical observations. For instance, let us recall Mr. I, the artist who lost color vision while retaining other visual capacities (see chapter 3; Sacks, 1996). He presents us with the case of a “Mary-in-reverse,” growing up as he did in a colorful world and gaining a broad, general knowledge relating colors to physical phenomena such as effects of mixing paint. Because of an accident, he then completely *lost* color vision while retaining the same knowledge as he had before. Whereas Mary gained *new* experiences when she stepped outside her gray world, Mr. I. could no longer see colors in the full presence of his body of color-associated knowledge. It would be extremely hard to maintain that studying neurophysiology and optics would have brought back his color vision.

11.7 Chalmers’s Double-Aspect Principle of Information

We already encountered the work of David Chalmers when reviewing the “easy” and “hard” problems of consciousness and his critique on neuroscientific reductionism. To recapitulate, he rejected the potential “neural correlates of consciousness” that neuroscientists came up with because these processes would only address “easy,” mechanistic parts of the problem (such as gamma oscillations accounting for perceptual binding of features) but leave the “hard problem” unsolved. My response, given in chapter 10, was that trying to identify *direct* correlates of conscious experience at a neural level is relatively fruitless because phenomena at a low functional–computational level do not directly illuminate how high-level processes correspond to full-blown consciousness.

Following up on the validity of Jackson’s knowledge argument, let us consider how Chalmers’s own account of consciousness fares relative to the neural and other theories he criticizes. First, Chalmers adopts a nonreductive approach, which agrees with the heuristic route (nonreductive but explanatory) I laid out above. Next, Chalmers (1995) takes experience to be “a fundamental feature of the world, alongside mass, charge and space-time” (p. 210). Recognizing that this approach supposes other properties above those postulated by physics, he qualifies it as a variety of dualism—“but it is an innocent version of dualism, entirely compatible with the scientific view of the world” (p. 210). The approach

is elaborated by identifying “bridging principles” explaining how experience arises from physical processes. A first principle refers to the coherence between the structure of consciousness and the structure of awareness. By “awareness” Chalmers means various functional phenomena we can associate with consciousness, such as binding, attention, global control over information processing—the kind of processes cast earlier under “easy problems.” One may notice a resemblance between this principle and the idea of isomorphism proposed by the Gestalt psychologist Köhler (1940; see also chapter 10). Briefly, structural coherence means that the hard and easy problems are tightly interconnected to each other, that consciousness does not “float free” relative to identifiable cognitive and computational mechanisms.

A second principle postulates organizational invariance, holding that two systems with the same fine-grained functional organization will have qualitatively identical experiences. “What matters for the emergence of experience is not the specific physical makeup of a system, but the abstract pattern of causal interaction between its components,” according to Chalmers (1995). This latter principle appears to amount to the familiar notion of multiple realizability.

Finally, Chalmers proposes a double-aspect principle of information. Inspired by Shannon’s (1948) information theory and Wheeler’s (1990) “it from bit” doctrine that posits a fundamental role of information across the entire domain of physics, Chalmers observes that information is key to both physical and psychological systems alike. Both types of system can be seen as changing their degree of uncertainty when their states are changing because of interaction or communication with some other system. The double-aspect principle holds that there is “direct isomorphism between certain physically embodied information spaces and certain phenomenal (or experiential) information spaces.”

Does Chalmers’s approach hold promise to solve some of the key problems of consciousness? As Shannon’s notion of information is based on statistical principles such as entropy and mutual dependence between system elements (see chapter 5), a caveat is that any system in nature, whether an exploding supernova or a melting glacier, can be claimed to be processing or transmitting information. This would imply that all active systems would have a phenomenal aspect as well, leading us back to panpsychism. Chalmers acknowledges this problem: “one possibility is that we need a further constraint on the fundamental theory, indicating just what sort of information has a phenomenal aspect.” Considering the strong evidence for unconsciously versus consciously representing brain systems and their awake and sleeping states (see chapters 6 and 7), I have argued that this should not be a mere possibility, but rather a necessity for a good theory—and that there is no single further constraint, but a whole list of requirements we need to work on (see chapters 8 and 9). Innocent as Chalmers’s version of dualism may be, it fails to offer a solution to the hard problem and to add new insights that would justify a revival of dualism.

A potentially fruitful way forward would be to radically modify information theory by incorporating semantics. Shannon’s information is tremendously useful for quantifying the

syntactical, statistical regularities relevant for mind–brain operations, but other, nonclassical approaches will be needed to understand problems related to meaning and phenomenal content.

11.8 Searle's Direct Realism versus Representationalism

Searle's Chinese room argument has become famous as an attack against functionalism and “hard” AI—the branch of AI claiming that mental processes can be emulated by computer programs (Newell, 1990). Astute and resilient as the argument has proven to be, we did face the problem that our own brains are also living in a kind of Chinese room (see figure 4.11). Your brain, too, is locked inside the chamber of your skull, and all it receives from the world are spikes which it needs to make sense of. Miraculously, this works—but how?

Let us first contemplate how Searle approaches this problem. In a view he labels “biological naturalism,” Searle (2000) casts consciousness as a biological phenomenon like any other. Conscious states are caused by neurobiological processes in the brain, analogous to the way digestion of food is caused by enzymatically catalyzed breakdown of macromolecules in our gastrointestinal tract. But, he adds, consciousness also has unique peculiarities: it has a subjective, unified nature and is set in a first-person perspective. Nonetheless, consciousness is amenable to objective scientific study. This study should obey dualism nor materialism for reasons already discussed. Searle's framework emphasizes the biological unity of the mind–brain, which demarcates his position from functionalist views. Although Searle (2004) has expressed support for the conceptual feasibility of zombies, he has also adopted a more liberal attitude toward multiple realizability in other work (Searle, 2000):

I believe [...] that understanding the nature of consciousness crucially requires understanding how brain processes cause and realize consciousness. Perhaps when we understand how brains do that, we can build conscious artifacts using some nonbiological materials that duplicate, and not merely simulate, the causal powers that brains have. (pp. 576–577)

Indeed, if one would persist that multiple realizability is impossible in principle, one would be forced to speak of brain properties that computer chips or other emulating devices could never have—ending up in a deadlock position similar to how vitalism got stuck in the late nineteenth century.⁵ Nonetheless, Searle argues that a conscious neural system requires causal powers beyond the capacities of today's computers and the physiological and structural principles we currently know. Searle does not specify what the nature of the “causal powers” that he uniquely attributes to the brain may be. However, if these “powers” would be organizational in nature, his way of thinking would be in line with the current multilevel account.

Coming back to the brain's cuneiform room, it is interesting to follow Searle's thinking about sense organs and perception. He opposes the “sense data theory” (Russell, 1912; Ayer, 1953) which has its roots in the work of, for example, Locke, Hume, Berkeley, and

Kant. This theory states that what we see or hear are not objects or events as they exist independently from ourselves in the external world, but the appearances, impressions, or representations we form via our sense organs and brain operations. A key argument against this theory holds that we are not actually *looking* at our impressions but at objects located in the outside world (cf. Searle, 2004). However, if one were to insist that *all* we can perceive is located in the outside world, it would become difficult to explain phenomena such as illusions or a visual afterimage you perceive when you first look straight at the sun and then close your eyes.

To deal with this problem, yet avoid the problems of Russellian sense data, Anscombe (1965), Hintikka (1969), Shoemaker (1990), Tye (1994), Dretske (1995), and others advanced *representationalism*. This alternative school of thought is generally positioned within the materialist tradition but at the same time attempts to accommodate, rather than deny or ignore, qualia. It is not a mind–brain theory in the sense of a comprehensive framework that tries to explain how representations arise from neural activity patterns (Dretske, 2003). When we experience snakes as rotating (see figure 8.6, plate 10), representationalists argue that we are not looking at an experience (or impression) that rotates but attribute the rotating quality to the *object* that is represented (a snake shape). Qualia are thus cast as represented properties belonging to objects that are also represented within (or by) the same experience. In other words, qualitative properties are not “stand-alone” items in phenomenal experience, detached from anything else in our experienced world, but are *about* something in it. This brings us back to the concept of intentionality: the rotational movement is an intentional content, that is, a property referring to something that is in this case unveridical, but is nonetheless represented. In sum, both illusory and veridical rotations are qualitative, intentional objects, but in the illusory case this object is nonactual whereas a veridical rotation is actual.

Searle defends an alternative to representationalism that he calls a “direct realism” or “naïve realism,” stating that it makes no sense to say that we “see” a sense impression. “If I extend my arm forward, I see my hand in front of my face,” Searle argues, and thus explains that what we perceive pertains directly to our physical bodies and their material surroundings (see also Austin, 1962). Representational or sense-data theories, he continues, are disastrous because they make it “impossible to give a true account of how human beings and other animals relate to the real world” (Searle, 2004).

Given that Searle advocates the causal powers of the brain as uniquely equipped for generating consciousness, how does he defend naïve realism? Searle reasons that a causal account of how we come to see the world—via physical stimulation, sensory receptors triggering impulse trains traveling to the brain, and so on—does not refute the idea that we actually do see the real world. The causal account does not prove that what we see is actually a “sense impression”—a representation derived from the real world via our senses. Searle also attacks the representationalist argument from illusion, stating that naïve realism must be false because it cannot account for perception of illusory objects or

hallucinations. When gazing at the yellow-blue circles of figure 8.6 (plate 10), we visually perceive a rotational movement whereas verification by other means demonstrates the circles are completely stationary. If our brains concoct such nonveridical constructs, so the representationalist argument goes, we should admit that veridical experiences are also constructs forged by the brain. Searle's counterargument is a linguistic one: in the case of illusions, we do not really "see" an object such as Kanizsa's triangle the same way we see a veridical triangle. Precisely because we acknowledge that Kanizsa's triangle is illusory, we should maintain a distinction between "seeing" and "having an illusion." When we perceive an object (that is, a nonillusory thing), we do not see "its appearance," because there is no way of segregating seeing this versus seeing the object itself, out there in the world.

Searle's arguments clarify the way we ought to *talk* about perception but do not solve the brain's cuneiform room problem. When I look out the window and see the orange roof of a neighbor's house across the street, the verb "to see" is not properly applied as meaning that, on one side, I am here and am looking at my brain's own representation of that roof, standing at the opposite side of the scene. The expression "I see an orange roof" does not merely refer to a single perceptual representation, because it invokes the notion of self-consciousness. The orange roof across the street is primary to my conscious experience. But, contrary to Searle, this conscious experience is entirely generated by the brain, as it is fed by streams of impulses from the retina. The constructive nature of conscious representation (see chapters 8 and 9) implies that we do not locate the object of experience inside our own head but rather in the spatial context of other elements of experience—the sky, the ground, the trees surrounding the orange roof.⁶ Indeed it is incorrect to say that I am looking at the brain's representation of all this, not because the representation is not there or would be irrelevant to perception, but because the act of "looking at" cannot be properly applied to something that is physically happening inside the brain—it pertains to whatever is "projected" by the brain as the current visual world. Seeing and looking are acts taking part in the projectional process. The notion of the (self-conscious) "I see X" is secondary and comes only after "seeing X." It applies to the object which is (really!) located in the world in relation to the "I" or the viewer, who also represents himself or herself as being positioned in that world. This self-positioning results from the configuration by which proprioceptive, kinesthetic, tactile, and visual cues are integrated to provide evidence as to where our own body is in space and how it is oriented (see chapters 8 and 9).

Classic experiments intervening with perceptual brain activity illustrate where Searle's argument goes awry. By electrically stimulating the temporal lobe of epileptic patients, Wilder Penfield evoked reportable subjective experiences such as the smell of burning toast (cf. Penfield & Rasmussen, 1950; Penfield, 1958). This smell was experienced by the patient as veridical although later on the patient understood that it resulted from the artificial stimulation in the physical absence of burning toast. The smell of toast evoked by brain stimulation is as much a genuine percept as the smell of toast that is physically burning in the subject's kitchen. It is only by (often post hoc) recognition of the context

that we come to recognize the experience as “illusory” or artificially evoked. Following Searle, there is no ground for fundamentally distinguishing an “apparent” from a “real” smell, but this does not refute that the stimulation triggered a brain process corresponding to the experience of burning toast. The same holds for phosphene experiences evoked by TMS pulses applied to the scalp overlying visual cortical areas (Cowey & Walsh, 2000; Antal et al., 2004). All experience is forged by the brain, and the brain has no direct, ghost-like access to physical objects external to the body—an access circumventing the route of sensory nerve fibers.

Neuroscience and neuropsychology abound with examples demonstrating that we do have illusory percepts that are experienced as veridical and are not distinguishable from “real” percepts often until well after the fact. Illusory experiences can be dramatically persistent. A schizophrenic patient may hear voices but will not recognize the self-generated origin of these sounds (see chapter 8). A patient suffering from Anton–Babinski syndrome will not recognize his or her own blindness, persevering in the assumption that he is still seeing (visual anosognosia; Eby et al., 2012). More commonly, in dreams we experience short, emotional and often bizarre episodes that we only acknowledge as counterfactual by the time we wake up.

In short, the evidence that the brain’s machinery generates representations we consciously experience is tremendously strong and undeniable. One may dryly reply, “Tell me what you like, but I still see my hand right in front of me.” No arguing with that, but this statement is also shorthand for saying your brain generates a higher-order representation that couples the visual representation of a hand to a multimodal representation of your own body in a common space in which both hand and body are represented. Altogether, Searle’s direct realism becomes harder to defend when realizing that it recognizes “mental properties” or other features that zombies cannot have (Searle, 2004) but also argues that we directly perceive material, external objects. How the two connect in this view remains unclear. In a pessimistic scenario, Searle’s view amounts to cryptodualism. On a more positive note, it encourages us to flesh out what the brain’s “causal powers” are.

Returning to representationalism as an alternative to naive realism, the current framework appears to be akin to this position at first glance. Of the various flavors of representationalism developed over the past decades, the “strong” variant defended by Dretske (1995, 2003), Tye (1994, 1995, 2000), and Lycan (1996, 2001) is particularly interesting for this comparison. To represent a certain quality, according to this variant, a simple representation (like the pixels of a digital photograph) is not enough to represent that quality; a specific *kind* of representation is needed, its requirements being defined in functionalist or other materialist terms. So far, the current account conforms to this theory. However, strong representationalism also follows Putnam (1975a) in taking representational content to be “wide” in that qualia are considered to be (veridical or illusory) *environmental* properties and are thus not purely determined by representations internal to the nervous system.

Two molecularly identical twins would experience different representational contents. This brings this variant close to externalism, which was argued against above.

Adherents of strong representationalism or externalism often refer to Putnam's (1975a) Twin Earth argument (see Burge, 1979, for a related argument). Assume that your brain was completely reproduced, molecule by molecule, and that the duplicate brain was transported to a Twin Earth, while your original brain resides where it is now. Twin Earth is a copy of Earth, except that water (H_2O) has been replaced by a substance that has exactly the same properties as water, and behaves the same, but has a different chemical identity (XYZ). Because the twin brain, when confronted with XYZ, will think of (or experience) this substance as water, it is representing XYZ *incorrectly*. The representations in both brains would refer to different objects and have a different intentionality although these brains are molecularly identical. Thus, a complete representation must include environmental properties.

The weakness of the argument is that the twin brain, situated on Twin Earth, has no way of knowing that it "misrepresents" XYZ as "water." The assumed "incorrectness" of the twin brain's representation has no relevance whatsoever to the way XYZ feels to the twin brain. All qualities of XYZ the twin brain experiences are the same as it would experience for H_2O on Earth. If the claim is that you have to know exactly what is happening in the outside world beyond what can be extracted from brain-generated processes ("wide content") to forge a complete and fully "correct" representation, the question remains how any individual could know all that. All information about qualities of all discrete, external objects, and all of their further environments, reaches us by sensory relays from periphery to brain, and if two molecularly identical twins were to receive the exact same sensory inputs, they would have no way of knowing that XYZ would be any different from water, and thus sustain the same experience. Once they start talking (or believing), one twin uttering "water" and the other "XYZ," this difference in verbalization already implies their brains *cannot* be identical. The utterings can only be different if they correspond to diverging brain states. Thus, the current theory is much in favor of "narrow" representationalism, which holds that molecular duplicates share the same qualitative experience (see, e.g., Shoemaker, 1994; Horgan, 2000; Levine, 2003; Chalmers, 2004).

11.9 Representations as Internally Generated Phenomena; Solipsism and the Public Domain

Searle (2004) and others before him also brought up the "public language" argument against sense-data theory and representationalism. This argument supposes that, in order to communicate successfully, people must have access to the same set of common objects of reference. When you and I are talking about "this table" in a conversation, we must be referring to the same object for this reference to make sense. However, is this necessary to communicate successfully? If you tell me, "You should not eat this apple; it is rotten" and point with

your finger to one particular apple among many, my brain will integrate this auditory input with what it perceives visually. I will couple the spoken word “rotten” (and its associations with bad taste, etc.) to a visual object with a particular complexion and shape, and hence this multimodal input instructs me to adopt an action policy (refrain from biting the apple). That your verbal communication contributes to my actions ensuring my well-being proves the success of this communication. However, nothing in this communication requires that you and I refer to the exact same, publicly observable phenomenon, or same percept. What looks like a brown–orange spot to you could look different to me. When we talk about “understanding each other” by way of public language, we refer to knowing how to apply another person’s verbally expressed knowledge to adjust our own behavior, not to knowing the other person’s subjective experiences. Successful communication does not require that our internal representations are identical or even similar. What is required, first, is a consistent coupling of auditory information to inputs from other modalities, such that the correlation is consistent over time (cross-modal consistency). One and the same color will not be given different names over time, and if an apple turns from fresh to rotten, its color description should shift according to the constraints of perceptual color space (see figure 10.4). Second, we should be able to generalize across individual instances of “rotten” and thus build abstract vocabularies, replacing references to specific objects or properties with categories thereof, and so on. My argument here is that people can communicate as successfully as they do in daily practice and yet experience only the representational patterns their own brains sustain—privately. Many philosophers may despise a defense of representations being private, but I do not see a fundamental, unsolvable problem here.

When we contemplate an object such as a green lawn, there is also no objective way to figure out what the “green” looks like “out there,” that is, independent of anybody’s perception. Theoretically, the percept of a neat green lawn might even be caused by tiny little insects crawling around in hollow eye sockets and pressing on mechanoreceptors, instead of photons hitting cones in the retina. The behavior of these six-legged creatures is just so accurate that they tap on the right buttons at the right times, in a well-orchestrated dance. One might object, “I can easily pick one of those bugs out of my eye socket and feel it by touch.” Alas, we are prisoners of a multimodal devil, because our tactile sense is also utterly controlled by other insects pressing on different mechanoreceptors at the right time, raising the illusion that what you actually touch is not an insect, but the curvature of your cornea. This Cartesian devil is also in charge of the sensory feedback you receive from your own actions. Gruesome as this metaphor may be—if the insect version of reality were true, you would be lucky enough not to notice the difference from “normal” reality. You would be able to enjoy the taste of chocolate and the sight of a Whistler painting just the same as you do now, remaining unaware of the insects as agents causing your sensations. The only way to go firmly beyond the idea of multimodal demons and obtain clues as to what external reality “really is like” lies in the construction of *abstract* world models.

This does not render our own, conscious experiences worthless when dealing with reality. On the contrary, they are the best possible estimate of what we are dealing with, both inside and outside our bodies. The argument is that expressions such as “looking like,” “resembling,” and “reflecting” become inappropriate when trying to compare our brain representations *directly* to whatever it is that activates our sense organs.

As far as we can gain knowledge about the nature of a viewer-independent, material world, physical models have laid out the best heuristics so far, using both experimentation and abstract, mathematical representation. A fork on a table may be subjectively experienced as “solid,” but according to physics such a feeling is better understood by electrostatic interactions between the object and our skin. Crimson red and shiny as the cape in the Velázquez painting of Pope Innocent X may be, this impression says little about the underlying dualistic nature of light, behaving both as an electromagnetic wave and a photon particle. One consolation physics has to offer when we face the epistemological gap toward knowing external reality is that its rigorous methods confirm that this reality deviates on major points from how we experience it subjectively.

As a neurobiological variant of “narrow” representationalism, the current theory bears the risk of being attacked as resulting in idealism. As best known historical representative of idealism George Berkeley (1707) argued, we can only know qualitative and quantitative properties of objects we perceive by way of ideas in our mind—so the only “real things” are the phenomena in our conscious experience. “Esse est percipi,” Berkeley said—“To be is to be observed.” We may assume a reality external to the ideas in our mind, but a pure idealist would deny that there is solid ground for this assumption. One of the many arguments against idealism is that it leads to solipsism, the view that only the I, one’s own mind, exists. It doubts, or even denies, the reality not only of an external world but also of other people’s minds, and it has trouble explaining how we can meaningfully communicate with other people.

Although the current framework is founded on internally generated representations, corresponding to a phenomenology that is essentially private, it differs from idealism in that it does allow for communication about people’s thoughts, abstract ideas, and life events. People can influence each other’s behavior perfectly well, as we construct models of people in general, and of specific persons we are familiar with—including models of how they feel and react to events. There is no fundamental problem because you and the other person’s brain can each rely on the cross-modal and temporal consistency of perceptual representations, but also on abstracting capacities. Using these, we build elaborate conceptual constructs of how external reality is organized, including other people’s minds and behaviors.

For example, could the reality-by-insect-dancing hypothesis explain the movement of the celestial bodies relative to our viewpoint on Earth? In principle, one might suppose that the tapping insects possess tremendously complicated motor programs in their nervous systems that could reproduce a pattern of retinal activation reflecting the way the moon, sun, and planets seem to orbit around the Earth. However, this would amount to a scheme

going beyond the complications and ad hoc assumptions of Ptolemy's astronomical theory and would thus be easily defeated by Newton's simpler and very accurate laws (Kuhn, 1962). The insects would have to "know" so much about the complexities of celestial movements over large periods of time that they would have to be God-like, omniscient creatures. In short, replicable observation, experimentation, and abstract model-building lead the way against solipsism, but let us not forget perception's memento mori that we cannot get completely rid of private, subjective aspects of experience—however innocuous in our mental and social functioning.

Physics may set an example in consciousness research in yet another way. Going back to the multilevel scheme of figure 10.2, we assume there is a gradual transition from neurobiological processes to mental processes—although we cannot imagine how this transition works exactly. There are many examples of such unimaginable, abstract steps in physics. Apart from the wave–quantum paradox of light, physicists have to reconcile quantum-mechanical uncertainty with macroscopic certainty, or with the likely existence of a "ghost-like" entanglement of two photons influencing each other across large distances, behaving as a nonseparable object (the Einstein–Podolsky–Rosen paradox; Aspect, 1999). Physicists excel at inventing new concepts and ways of thinking when confronted with problems that cannot be solved within an established framework of assumptions. As classical elementary particle physics failed to account for all properties of hadrons, Gell-Mann, Feynman, and others proposed a quantum chromodynamics in which these particles were assigned a flavor or color in addition to properties such as mass, velocity, and electric charge. Scientists will have to be similarly creative and flexible as mind–brain theories will be unfolding further, accepting that some conceptual steps may remain paradoxical and irreducible. It is a comforting thought that a child's developing brain may also be regarded as a natural "experiential physicist" in that it synthesizes its own perceptual qualities and cognitive categories.

11.10 Animal Consciousness?

The ant *Camponotus leonardi* lives in the tropical rainforests of Thailand and starts behaving like a zombie when infected with *Ophiocordyceps* fungi (Hughes et al., 2011). Healthy exemplars live in tree canopy nests and are rarely found on the ground. The fungus, however, prospers best in the moist and warm environment of the undergrowth. Once infected, a ball-on-a-stick-like mushroom starts growing out of their head, and the ants descend to the ground where they aimlessly wander about in tortuous paths through the under-story plants (see figure 11.1, plate 13). Their jaws clasp a juicy vein of a leaflet, and there they slowly die. Since the onset of infection, the fungus has been proliferating single cells into the ant's head, and the substances secreted by the fungus have somehow taken control over its behavior to subserve the interests of the fungus. The sad case of *Camponotus* does not prove that healthy ants would be conscious or can behave like zombies in the

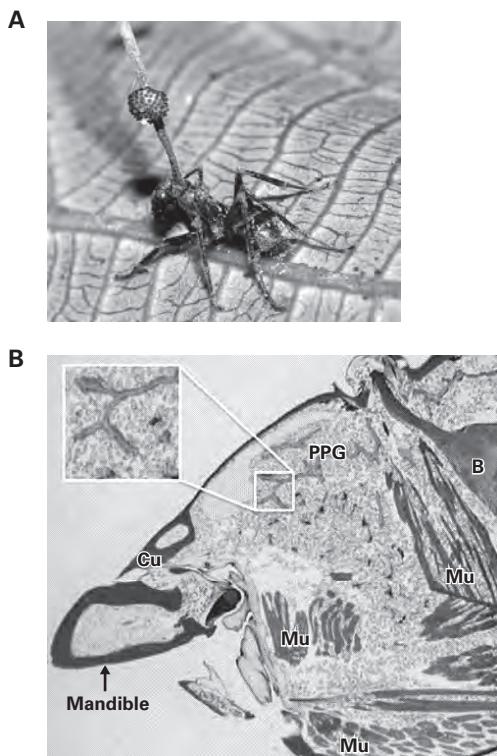


Figure 11.1 (plate 13)

Fungus infection of the ant *Camponotus leonardi* illustrates the debate on zombies and animal consciousness. (A) Once the fungus *Ophiocordyceps* has infected the head of the ant, a stalk grows from its back to produce a fruiting body from which spores will be propulsively released. (B) Sagittal section from the head of an ant performing a “death grip” in a leaflet at the moment of fixation. Fungal hyphens are represented in the zoom-in of space around the postpharyngeal gland (PPG) and as small gray blobs filling the head and mandibles. B, brain; Mu, muscles; Cu, cuticle. The erratic random-walk behavior and the death grip of the ants can be interpreted as a form of externally induced, fungal control over the animal’s behavior, effectively reducing it to a living “zombie.” We can conceive of this behavior as happening without an “inner,” mental life, which however does not imply that mental states are separable from physical states of the nervous system. Arguments relying on conceivability or imagination suffer from the same weakness when considering human or lower-animal versions of zombies. From Hughes et al. (2011).

philosophical sense. However, it does illustrate the debate on animal consciousness and the limited evidence derived from purely behavioral observations.

What rational arguments do we have to claim that animals are conscious or not—and could their consciousness, if existent, be graded across species? Complexity of behavior, including communication, may serve as a first proxy to the richness of sensory, emotional, motivational, and motor planning processes in the brains of animals, but one of the main tenets of the current account has been to link consciousness to predominantly sensory

representations because of the many cases in which it can be dissociated from behavior. As a second proxy to consciousness, we may consider the anatomical or physiological similarity between the human brain and that of another species. Beings having similar brains are more likely to sustain similar experiences. A first disadvantage of this approach is that it does not yield a principled, insightful way to study consciousness across species and leads to oversimplifications, such as “a bigger neocortex means more consciousness.” Secondly, relying on similarities to the human brain neglects the possibility that phylogenetically remote species, such as octopuses, may have developed very different, nonhomologous mechanisms to generate consciousness.

A third approach for assessing animal consciousness is to ask to what extent the requirements for conscious representations are met by the nervous systems under scrutiny. Recapping chapters 8 through 10, to qualify for full-blown consciousness, systems are at least required to integrate different sensory (or imagined, retrieved) multimodal inputs, so that these inputs can be identified as representing a feature in a specific modality. In this process of integration, features are represented as being part of coherent objects which are, in turn, embedded in immersive situational sketches and are attributed further meaning by way of mnemonic and interpretive mechanisms. A further requirement is to allow for both temporal stability and flexibility of percepts. These core requirements are realized in a hierarchy of low-level processes of representation, at neural and ensemble levels, up to the higher levels of meta-networks. In addition, we paid attention to a less essential requirement, namely, the alignment of spatially mapped information, such that percepts become unitary and externally represented, creating a first-person situational positioning and perspective.

As long as general neural mechanisms meeting these requirements have not been identified, it is impossible to pinpoint exactly which species may fulfill these requirements. However, the requirements do allow us to infer that at least some systems properties should be present. For modality identification and constructing enriched situational representations, a brain will need profusely interconnected topographic maps set in multiple modalities, multimodal convergence zones, or both. To enable phenomenal and cognitive interpretation, we postulated echoing processes between sensory–sensory and higher-order areas (see figure 9.10). For temporary stability, we referred to networks with continuous attractor properties and so on. Using these requirements and their putative mechanisms, we may attempt to make limited predictions on consciousness in animals.

All key and auxiliary requirements can be graded in nature. This gradual nature stems from the realization that situational representations can be graded in unimodal complexity (e.g., considering only visual motion in a scene), multimodal complexity (enrichment derived from a number of modalities being aligned in a situation; see figures 9.6 and 10.2), complexity or “depth” of interpreting situational elements by their relational networks, and finally, in salience or intensity (related to attentiveness). When I go swimming in the North Sea, a jellyfish may lurk beneath the water surface until it stings when I accidentally bump into it. Could it somehow be consciously doing so? This is improbable if we

take “conscious” as referring to the kind of state we have been discussing. Although some cnidarians, such as box jellyfish, are recognized to have more complex nervous systems than simple, diffuse “nerve nets” (Satterlie, 2011), there are many species in which sensor cells are thought to convey information about environmental variables such as light intensity directly to neuromuscular effectors, which thus serve as central convergence points of sensory integration. Scyphozoan jellyfish, for instance, possess organs known as rhopalia containing light sensors (ocelli) and statocysts for maintaining balance, and these sensing signals converge directly on the motor components of these organs, subserving the pacing of swimming contractions. Unlike in the parietal and temporal cortices, a neural integration process for identifying and comparing different sensory streams in a spatial, quasi-stable representation appears to be lacking. Leeches may present a more advanced case as they integrate chemosensory, thermal, visual, tactile, and proprioceptive signals to decide between feeding and swimming behavior (Harley et al., 2011; Gaudry & Kristan, 2012). However, what makes this case different from the complex, multimodal integration meant here is that widespread, bidirectional communication between topographic maps of different modalities is probably lacking (cf. figures 6.3 and 6.11). In leeches, the integration is of a feedforward nature, directed at biasing the central pattern generators of the leech motor system (Gaudry & Kristan, 2012).

Several steps up the scale of evolutionary complexity, insects and squids exhibit complex behaviors, mediated by well-developed nervous systems. Odor detection in locusts, for instance, is accomplished by serially chained networks of cells in the antennal lobe, mushroom body, and other ganglion cells (Laurent et al., 1998; figure 11.2, plate 14), analyzing olfactory information along progressive stages of processing. The function of mushroom bodies is generally related to learning and multisensory integration. Projections from this structure back to primary olfactory sensors in the antennal lobe provide a top-down control potentially supporting filtering and categorization (Hu et al., 2010). However, the integration of olfactory, visual, and mechanosensory information is not limited to the mushroom bodies and involves an intricate web of feedforward, feedback, and lateral interactions involving multiple structures (Duistermars & Frye, 2010). The multimodal richness of insect sensory processing may not hinge on a shared spatial topology, and the means for interpretive feedback are probably less than in mammals, but this class of arthropods does present a case of intermediate complexity for which it is hard to completely exclude some primitive kind of conscious representation. On the one hand, the basic hardware for multimodal representation seems to be present. On the other, we have seen that many subcortical systems in the mammalian brain do not qualify for conscious representation, and insect nervous systems cannot be claimed to surpass these systems in complexity. Recalling the argument from nonhomologous mechanisms, we have to leave the case of evertebrate consciousness undecided for now. A similar case can be made for the octopus.

Can we say more than that it is hard to exclude certain species from having consciousness? The answer I would like to drive home is, in principle, yes—but one has to be willing

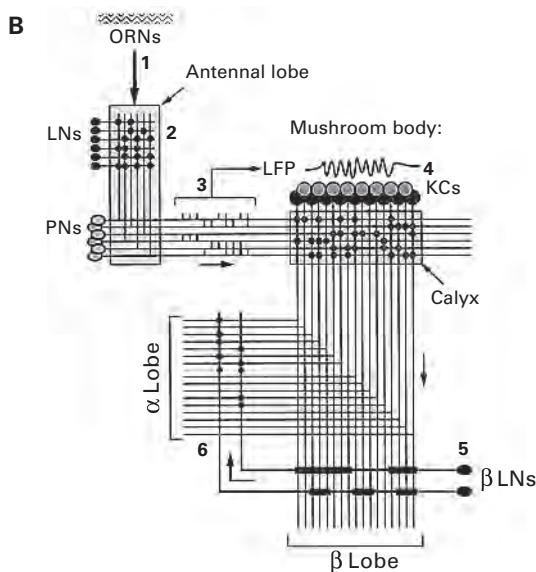
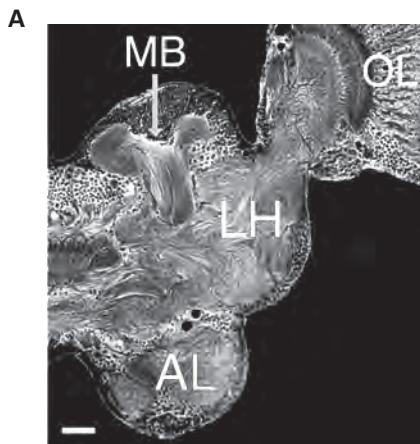


Figure 11.2 (plate 14)

Networks for sensory processing, multisensory integration, and memory in the locust brain. (A) Transverse section of the locust brain (Bodian stain), showing the antennal lobe (AL), mushroom body (MB), lateral horn (LH), and optic lobe (OL). Calibration: 80 μm . (B) Chemosensory signals are relayed from olfactory receptors neurons (ORNs) on the antenna to the antennal lobe. This input is processed by projection and local neurons (PNs and LNs). Projection neurons transfer their spike signals to the mushroom body, where Kenyon cells (KCs) are rhythmically activated. Synchronized excitatory postsynaptic potentials (EPSPs) elicited by this PN input give rise to 20–30 Hz local field potential (LFP) oscillations in the calyx. In turn, Kenyon cells modify activity subsequently in the beta and alpha lobe. Several top-down feedback and multisensory input pathways are not included in this diagram (but see Hu et al., 2010, and Duistermars & Frye, 2010). The locust brain presents a puzzling case of intermediate complexity for which it is difficult to exclude the existence of basic, primordial conscious representation. It shares important functional features with corticothalamic systems of the mammalian brain, including top-down recurrent projections and multimodal integration, which are however also found in basic form in subcortical structures of the mammalian brain not implicated in conscious processing. (A) from Perez-Orive et al. (2002). Reprinted with permission from AAAS. (B) from Laurent et al. (1998). Reprinted with permission from Cold Spring Harbor Laboratory Press.

to thoroughly investigate nervous systems based on requirements set in advance. Without this, it will be impossible to decide whether, for instance, amphibians or reptiles are more likely to possess consciousness than highly evolved arthropods. Like these evertebrates, reptiles and amphibians present an intermediate case. Their spinal cord, brain stem, hypothalamus, cerebellum, tectum, and basal ganglia are roughly comparable to those of mammals (Marin et al., 1998; Stephenson-Jones et al., 2011), but the lack of a clearly developed neocortex (Striedter, 2006) makes one wonder whether any form of consciousness would be possible. Loss of neocortical function in humans is accompanied by loss of consciousness, but is this also the case when you have no neocortex to begin with? Theoretically, it seems possible that multimodal-representational functions could be mediated by subcortical brain structures if no cortex has ever developed to do the job.⁷ Reptiles and amphibians possess a tectum (i.e., the superior colliculus in mammals) where visual inputs converge with auditory and somatosensory inputs (Stein & Gaither, 1981; Marin et al., 1998). However, as in mammals, this structure functions primarily in directing eye movements, orientation responses, and defensive behaviors, not in sustaining sensory–cognitive, situational representations. A similar argument applies to the basal ganglia, which are well developed in amphibians and reptiles but probably serve the same nonconscious functions in motor control as seen in mammals (Marin et al., 1998; Stephenson-Jones et al., 2011; see chapter 6). At present, arguments to claim that the basal ganglia of amphibians and reptiles have a greater complexity than those in mammals are lacking.

All in all, animal consciousness remains a treacherous battleground from which it is hard to draw firm conclusions. To no small extent, this is also due to the lack of sophisticated behavioral tests (Weiskrantz, 1995; but see Stoerig & Cowey, 1997). Nonetheless, the problem will become more empirically tractable when investigations proceed to examine whether systems requirements are met in particular species, along with seeking evidence on homology and overt behavior that is indicative of conscious processing. In doing this research, we will gradually uncover more and better neural “markers” for consciousness—for instance, population-scale electrophysiological signatures of particular conscious states.

As a final point, let us return to Nagel’s bat. Not surprisingly, we still do not know what it feels like to be a bat. Will it feel like anything at all? When I read the fascinating paper of Dear et al. (1993) on cortical representations of acoustic scenes in the big brown bat, *Eptesicus*, I was persuaded that it would, in principle, be possible to make a statement on this. The paper describes how the bat’s tonotopic map transforms incoming sensory echo signals, arriving sequentially with different delays, into an accumulating representation of objects located at different distances within the environment. Following a biosonar vocalization, neurons respond very quickly (~10 ms) to the echo reflected from objects at short range. As echoes from more remote target objects are coming in slightly later (~22 ms), neurons start responding to these more distant targets, but the short range remains concurrently represented, and so on for greater distances. Overseeing this temporal-to-scenic transformation, one cannot help but notice how rich, apt, and smart this machinery

is in getting the job of environmental representation done. We cannot tell whether the bat “hears” or “sees” this representation, but this distinction is less relevant. Relevant is that the neurophysiology of the system is probably advanced enough to state that bats *do* have dynamic, integrative representations which—by inference—probably “feel like something.” Surely further research is needed to discover how other requirements such as multimodal integration are met, but this research is a step toward confirming that bats are sentient. Analogous to reverse engineering, we may identify the shimmer of a basic method here to infer “mental” processes from neural processes—an approach we may label “reverse psychology.” Given the representational capacities of a system, revealed by neurophysiology and psychophysics, we might reversely reconstruct what experiences will be like in such a system.

11.11 Computer and Robot Consciousness

The question of computer consciousness has much in common with the problem of animal consciousness but is set in a different light because some of the current supercomputers already display striking cognitive abilities in playing chess (Deep Blue) and language-based knowledge games such as Jeopardy (Ferrucci, 2010; cf. Ferrucci et al., 2013; http://researcher.watson.ibm.com/researcher/view_group.php?id=2099). These achievements relying on brute-force computation seem to contrast with the poor performance of robots in real-world environments, relative to what animals can do. Among the most advanced performances are a robot that uses vision to pick up novel objects from a dishwasher (e.g., Saxena et al., 2008) and a robot catching a ball that is thrown toward it (e.g., Birbach et al., 2011). The underlying computations have to be performed in real-time, and it is an astonishing feat that even these basic motor operations are carried out successfully. These robots have camera eyes, convert images into parsed scenes from which objects can be recognized, predict a ball’s future ballistic trajectory, and undertake accurate reaching and grasping actions within seconds. However, what do these supercomputers and robots tell us about consciousness?

Kurzweil (2005) has predicted that computers will surpass human intelligence by the year 2045, leaving us far behind as time proceeds toward the twenty-second century. His prediction is largely based on computing power (often referring to “Moore’s law,” holding that the number of transistors on an integrated circuit doubles about every 2 years), not on research examining how this power can be harnessed to enable smart, adaptive behavior, thinking, and conscious representation. Focusing on consciousness, it is clear that even the best current supercomputers do little in the way of fulfilling the systems requirements outlined earlier. Like desktop PCs, neural networks, or iPhones, they read inputs and perform massive computations on these but fail to identify what these inputs are about. We, as human users, instantly recognize what they are about, but inside the computer we find no systems mechanisms for modality identification, spatially organized representation,

and cognitive interpretation. This is also a principal objection against speculations that the Internet already is, or may ever become, conscious: there is no shortage of information exchange, but the system is utterly lacking the capacity for self-generated, high-level interpretation.

What about a computer's built-in mechanisms to identify pieces of information by way of memory addresses or to interpret inputs by way of look-up tables? The problem is that these mechanisms operate separately and do not ensure a rapid, flexible integrated process of representation, such as conceptualized by relational phase coding and dynamic echoing networks (see chapter 9). There is a simple reason why computers do not need to have all this. We built them simply to hand over the results of their computations back to us: it is the human user who has to make sense of the output. It would be a nuisance having a computer that does not return the raw results to us, but only its own, self-opinionated interpretation.

Despite their current limitations, the development of robots offers more hope when it comes to implementing some form of consciousness, however primitive. This hope is not inspired by their current computing power. It is because they face the same challenges animals face when trying to survive in outdoor environments: in the end they will need to construct quickly recognizable, multimodal, and flexible representations of the world around them—defined earlier as the core function of consciousness (see chapter 6). To what extent today's most sophisticated robots realize this function is a matter of debate. However, roboticists are at least implementing engineering solutions to grant artifacts the benefit of multisensory integration. For instance, a robot navigating an open space and searching for a target object benefits from the multisensory fusion of kinesthetic, visual, and echosonar cues, and the same applies to robots responding to auditory–visual human behavior (Arena et al., 2009; Trifa et al., 2007; cf. Wyss et al., 2006). As yet, this is far from robot consciousness, but the approach is promising because it allows one to combine neuromimetic mechanisms for high-level representation with the same ecological constraints animals face when navigating in real time.

When trying to decide whether human-made artifacts are intelligent and conscious, Turing's test provides an initial benchmark. A human interrogator sits behind a terminal and poses questions to a computer or human being connected to this terminal, invisible to the interrogator. In a question-and-answer session, the interrogator finds out whether he is dealing with a human or a machine. From a consciousness perspective, the problem of the Turing test is that false negatives as well as false positives may be found. An agent may be conscious yet unable to answer the questions because of limited linguistic or motor capacities. Watson, the supercomputing Jeopardy champ, may be smoothly answering many questions using Google-search-like strategies and sophisticated natural language processing routines yet be fully unconscious because it lacks appropriate representations. Encasing a supercomputer in an android robot with facial expressions that bridge the “uncanny valley” (Gray & Wegner, 2012) does not help, as behavioral markers are dissociable from representational capacities. On the positive side, the Turing test protects us from being

prejudiced by the appearance of the pretending agent. On the down side, it tempts AI programmers to invent tricks fooling the interrogator, preventing us from examining the pretender in detail and achieving real advances.

Koch and Tononi (2008) proposed a variant of Turing's idea: place the agent in front of a scene and require it to interpret the picture in terms of its overall meaning and the relationships between its elements, showing it got the gist of it. An image of a man pointing a gun at the agent should evoke a defensive response or an utterance like "Point that gun away from me!" This test offers additional value in assessing a system's ability not only to process visual input but also to analyze and make sense of it. The scene has to be parsed into objects, and objects have to be classified and recognized as having a behavioral and emotional significance. Also here, however, unconscious systems implementing sophisticated image analysis and classification-and-search routines may pass the test.

Therefore my plea comes down to more comprehensive testing, one that allows us to study the agent's behavior, vocal utterances, but also the complexity of its sensory-representational and interpretive systems as gauged by the anatomical, information-processing, and integration-related markers of system function at multiple levels. By reverse psychological engineering, we can investigate the system's representational capacities from the viewpoint of coding elements and larger aggregates and, hence, infer (together with behavioral observations) what the agent is capable of. With this complementary extension of the Turing test, future generations of robots are predicted to climb slowly and gradually up the ladder toward the complexity of human consciousness. Some robots may approach this pinnacle via the route of multilevel representation, others via an emphasis on cognitive interpretation, but in any case reaching this goal will entail a long and tortuous road and quite possibly take at least a full century to achieve.

11.12 Disorders of Consciousness and Future Neurotechnology

Facing clinical disorders such as comas and vegetative states makes one reserved toward claims concerning understanding or simulating consciousness. Nonetheless, some predictions can be made as to how current neuroscience may alter clinical practice in the long run. We are just now coming to grips with understanding if and when vegetative patients may engage in conscious processing. In an fMRI study, a vegetative patient was asked to alternately think of playing tennis or moving about in his house, generating activity changes in his premotor cortex and occipitoparietal junction region serving as "yes" and "no" answers when prompted to address questions (Monti et al., 2010; Fernandez-Espejo & Owen, 2013), and brain areas involved in these cognitive processes were correspondingly activated to indicate correct answers. Although it seems plausible that these patients are conscious but cannot express their thoughts by voice or gestures, it is also possible that they can sustain instructed information processing in these areas without consciousness. For instance, the neural processes corresponding to "playing tennis" as simulated motor

activity may occur in a rather isolated manner, not being integrated with visual and somatosensory imagery (Plum et al., 1998; Schiff et al., 2002). Despite this caveat, there is a fair chance that some vegetative patients are conscious but fully “locked-in,” which has deep legal and ethical implications. The mere possibility is already boosting clinical research to find ways to restore communication with patients, for example, by way of brain–machine interfaces (Guenther et al., 2009; Hochberg et al., 2012; Fernandez-Espejo & Owen, 2013).

Studies on disorders of consciousness are expected to stimulate the search for alternative, better neurophysiological markers of conscious states. Application of measures for neural integration to patients with diminished consciousness is beginning to offer alternative ways to assess consciousness besides classical behavioral markers or rough assessment of the scalp EEG (e.g., King et al., 2013). Casali et al. (2013) applied TMS pulses on the scalp of patients recovering from a coma (i.e., in a vegetative, minimally conscious or locked-in state) and studied the temporal and spatial spreading of the evoked neural response, gauged by an array of scalp EEG electrodes. Conscious states were marked by a significantly higher “perturbational complexity index”—reflecting the informational complexity of the evoked spatiotemporal EEG response—than in impaired states, including sedation or anesthesia. According to Tononi and Edelman’s hypothesis (see chapter 5), this higher complexity would indicate a higher degree of integrative processes that mediate consciousness, but this interpretation assumes the perturbational complexity measure accurately and critically reflects the neural mechanisms underlying consciousness, which remains to be verified. Nonetheless, further development of systems physiological markers, such as those based on Granger-causal interactions or graph theory (Seth, 2008; Barrett et al., 2012; Schiff et al., 2014), should spur on finer assessments of the exact state patients are in and can thus make a difference for deciding whether or not to prolong a patient’s life and how to continue treatment.

Looking farther ahead, we may ask how consciousness disorders may be more effectively treated. The definition of these disorders is broadly taken here, comprising not only coma and vegetative state but also schizophrenia and multiple-personality disorder. In a study by Schiff et al. (2007), a minimally conscious patient was treated with deep-brain stimulation of central thalamic nuclei, showing partial improvement of arousal level and motor behaviors such as limb movement and feeding. This method may be refined to a more personalized approach by designing a specific pattern of pulse trains that exactly matches the oscillatory properties of the patient’s thalamocortical circuits, which thus resonate to the stimulation pattern. We may foresee the implantation of miniaturized optogenetic devices and neurochips (Marblestone et al., 2013), capable of both stimulating brain tissue and recording from it, so that the electrophysiological tissue characteristics can be used to compute the optimal stimulation pattern in a closed-loop fashion. For truly chronic brain–machine interfacing, ultraminiature “neural dust” particles about 0.1 mm in size have already been proposed (Seo et al., 2013).

Earlier on, we visited the subject of schizophrenia in relation to distorted self- and reality representations. Many schizophrenic patients can be pharmacologically treated using an ever expanding range of drugs. A large fraction of schizophrenics, however, have no or little benefit from such drugs and continue to suffer from paranoid hallucinations and other symptoms such as conceptual disorganization. As florid hallucinations and poor reality monitoring have been associated with prefrontal dysfunction and activation of auditory cortex (Dierks et al., 1999; Vinogradov et al., 2008), a long-term effort may be undertaken to arrest this global pattern of aberrant brain activity by personalized electrophysiological devices intervening with such patterns only as they occur. This approach would avoid the great disadvantage of the persistent effects of drug treatment, often including desensitization. Combined with subjective reporting, these devices will be able to record and detect aberrant activity but also to apply electrical or optical pulse patterns to arrest hallucinations and restore cognitive control.

Our aging society poses problems which seem even harder to solve, such as Alzheimer's dementia. Here we observe a progressive decay of declarative memory, personality, attention, and perceptual abilities (cf. Wilson et al., 2012). Large-scale attempts are being made to unravel the molecular machinery that goes awry in this type of neurodegeneration, with the hope of finding a stick that can be pushed into the wheels to stop its progress (Haass & Selkoe, 2007; Querfurth & LaFerla, 2010). In a more pessimistic scenario, this decay cannot be stopped because it is caused by a multiplicity of molecular defects, resulting from an inability of neurons to repair themselves—that is, errors in their DNA, RNA, and protein production. In case genetic–pharmacological treatments fail, an alternative approach may be to progressively replace the decaying cellular machinery with neuromorphic devices that can be interfaced with intact living tissue (cf. Hampson et al., 2013). A first hurdle in this approach is the transfer of neural signals, recorded from the patient's brain, to a portable memory replacement device, but this problem is relatively tractable. Via a second neuron–silicon interface the patient would be enabled to self-generate commands for retrieving a memory appropriate for solving the query at hand. Here the engineering becomes very complex. Subsequently, a reactivated memory trace needs to be converted from its representation on a chip to electrical pulse patterns stimulating perceptual or cognitive areas of the patient's brain in exactly the right way. The vexing and currently unsolved problem here is to have the pulses hit the appropriate neural connections for the recall of neocortical information specifically associated with the trace. Large gaps between future fantasies and neuroengineering solutions remain to be crossed.

In summary, a bird's-eye view on philosophical schools of thought led us to demarcate the current theory from classical positions such as dualism, materialism, and functionalism. In addition to the dualist's problem of mind–brain communication and causality, it was noted that zombies might seem to be conceivable but are not realizable: the construction of the neural machinery appropriate for zombies would inevitably give rise to consciousness. Following Jackson's argument, the reality of phenomenal experience was acknowledged as

much as the reality of neural levels of processing, placing the theory away from eliminative materialism and classic functionalism. This situates the theory in an intermediate position that can be best described as a nonreductive, multilevel, neurobiological form of representationalism ('neurorepresentationalism'). Although representationalists have not been typically concerned with the problem of how neural aggregates give rise to consciousness, the "externalist" stream in this school is much more distant from the current view than the "narrow" stream emphasizing that representations are generated in and by the brain.

As such, the current discourse also better fits nonreductive physicalism than monism, where "physicalism" is grounded in a broad-sense, flexible physics that is willing to incorporate new concepts and terminology, as has been the case historically. In general, monism rejects the idea that mind and brain are different substances, or have fundamentally different properties. Instead, mental and physical phenomena are distinct manifestations of a single entity, a mind–brain unity (cf. Nagel, 1986). The positioning of the current account within physicalism relates to the close association of monism with panpsychism and to the asymmetry in mind–brain relationships noted earlier: mental phenomena depend on, and are determined by, physical phenomena and the presence of physical neurons—but conversely, it is *not* the case that physical phenomena are similarly determined by conscious minds. Neurons can exist in the absence of consciousness, and changes in the brain may well occur without mental changes (Kim, 1993; Van Gulick, 2001). A mind–brain unity is only realized when very particular systems requirements are met.

Let us conclude with expressing the hope that all of our theoretical and experimental efforts toward understanding the mind–brain will ultimately be of benefit to those who suffer from its dysfunction.

Notes

Chapter 1

1. Throughout this book, “consciousness” and “awareness” will be used interchangeably although “consciousness” will be used as the preferred term. Some authors have made a distinction between the two terms (e.g., Chalmers, 1996), but in this book I will treat them as referring to the same basic phenomenon.
2. Under this definition, subliminal perception would be a self-contradictory concept because it is taken to occur “below the threshold of consciousness.” Instead, this phenomenon can be better referred to as subliminal, or nonconscious, processing of inputs.

Chapter 2

1. Proprioception refers to sensory information derived from stretch receptors in the skeletal musculature, including tendons and joints. The brain uses this information to determine the relative positioning of, and tension on, body parts. Sometimes information on body motion and head orientation originating in the inner ear (vestibular system) is included in the overall sense of proprioception. Kinesthesia is a term often used interchangeably with proprioception but is more strongly connotative of muscular and body movement.
2. In the human brain, posterior parietal cortex is also referred to as Brodmann’s area 7. The neuroanatomist Korbinian Brodmann (1868–1918) divided the human cerebral cortex into as many as 52 distinct areas based on staining characteristics of histological sections, such as the thickness or cell density of particular layers (laminae) of the cortex.
3. As opposed to “neocortex,” denoting the convoluted parts making up most of the cerebral cortex in humans, the term “paleocortex” comprises the parahippocampal gyrus, piriform cortex, olfactory bulb, and other olfactory regions. It has only three to five layers, in contrast to the six layers of the neocortex. Next to neo- and paleocortex, the hippocampus is also considered a cortical structure and belongs to the “archicortex,” which is characterized by three or four layers.

Chapter 3

1. Different definitions of functionalism have been used. In a variant of the definition used here, multiple realizability is taken as the key element of this school of thought, whereas the role of mental states as causal intermediates is deemphasized.
2. It is currently debated whether the neural substrate of working memory primarily resides in the lateral prefrontal cortex or also involves more posterior regions such as parietal cortex and inferior temporal cortex (cf. Postle et al., 1999). Likewise it is somewhat uncertain whether lateral prefrontal lesions impair patients on delayed-response types of task because working memory or other cognitive functions (e.g., attentional control) are hampered (cf. Müller & Knight, 2006). To induce maximal working memory deficits, higher posterior regions, in addition to prefrontal cortex, would likely have to be inactivated (a manipulation that may well entail additional deficits in perception and input interpretation).

3. Here, attention should be paid to different definitions of consciousness used in the literature. Linkage of consciousness to language is often seen in definitions involving reflexive consciousness or self-consciousness, whereas this is regarded as a specific, higher-order form of consciousness in this book.

Chapter 4

1. Cited in Boring (1950, p. 82).
2. The term “unit” is used in the context of neural network models to denote a neuron or larger type of computational “node” such as a group of neurons performing one and the same function in the overall network.
3. Stereodisparity-sensitive neurons specifically fire when there is a discrepancy between the inputs from the left and right eye—that is, when they differ to at least a particular degree. This degree depends on the distance from the perceived object to the eyes. Such neurons are found in the primary visual cortex and at higher cortical levels of processing (Roe et al., 2007).
4. English translation adopted from Boring (1950).
5. Neural networks share this problem with traditional artificial intelligence (AI; e.g., Newell, 1990), which works with symbolic computation. Until now the field has not provided a satisfactory account of how computer systems may self-generate content or meaning associated with symbols (see also chapter 5).
6. From here on I will use the term modality identification problem because the labeled-lines hypothesis refers to a specific idea about the nature of nerves, sensory receptors, and nerve fibers, whereas the central problem is of a more general nature.

Chapter 5

1. Panpsychism refers to the notion that “psyche” (Greek for “soul” or “mind”) is present everywhere, all throughout (“pan”) nature.
2. Much of the dynamics of Hopfield nets indeed resembles the behavior of spin glasses, which helped to inspire the original Hopfield (1982) model.
3. Sleep stages are divided in various types, of which REM and non-REM sleep are most widely used. REM sleep is characterized by rapid eye movements and is often associated with dreaming. Non-REM sleep is subdivided into four stages, going from near-wakefulness (stage 1) to successively deeper stages up to stage 4. Stage 4 is characterized by high-amplitude, slow waves (0.5–4 Hz) in human scalp EEG and is often called “slow-wave sleep.”

Chapter 6

1. One may advance the counterargument that the cerebellum has a secret “conscious life” on its own, separate from thalamocortical systems. However, because we do not have access to such separate experiences, if existent, and would not experience their loss in the case of cerebellar damage, no further attention is paid here to this theoretical possibility.
2. Studying anencephalic infants, who are born largely without neocortex but with sophisticated subcortical circuitry, Merker (2007) adopted a different view on basal ganglia and other subcortical structures such as the superior colliculus, arguing that the emotionally expressive behavior found in these children points to consciousness in the absence of extensive neocortical tissue. This viewpoint is not adopted here because critical information regarding perceptual and reporting tests (e.g., on blindsight) is lacking.
3. Curiously, a subset of GABAergic inputs to cells in the suprachiasmatic nucleus turns excitatory during the subjective night phase of the animal (De Jeu & Pennartz, 2002), but this appears to be an exception to the general rule.
4. Cases of nonlocalized or “diffuse” sensations may, however, exist, such as a “Gemeingefühl” (general sensibility) or nonlocalized pain. See also chapters 8 and 9 for further discussion.
5. Although I argued against a strict dichotomy between access and phenomenal consciousness in chapter 5, it is nevertheless useful to maintain distinctions between the various aspects of consciousness, in particular behavioral reporting *versus* sensory-representational aspects. A functionalist objection to such distinctions may hold that representations per se are inseparable from some kind of behavioral expression because the latter is needed to obtain a report (e.g., verbal) about someone’s experience (cf. Cohen & Dennett, 2011). However,

sensory representations and behavioral reports may occur segregated in time, and, by virtue of memory, a behavioral report does not always have to follow immediately once a representation is generated. An example is given by a nightmare scenario in which patients are subjected to anesthetics for surgery and remain aware of specific details of the operation, while being unable to respond as it is ongoing (see chapter 7; for further discussion, see, e.g., Block, 2005, and chapter 11).

Chapter 7

1. Synchrony—originating from the greek *syn* (together) and *chronos* (time)—is the phenomenon that events occur “together in time”; their occurrence is temporally coordinated. A more specific use of the term refers to simultaneity of events in time, but here the broader meaning is adopted. In neurophysiology, synchronization is often observed between two or more periodic signals, for example, neurons bursting in unison or EEG traces recorded from the scalp. Phase synchronization is defined by oscillatory signals having a consistent phase relationship, that is, the relative phase angle of one signal with respect to the other is consistent over time and stays within a certain range of degrees. In the early days of EEG recordings, brain activity was labeled “synchronized” when voltage traces showed periodic, large-amplitude fluctuations, such as found in slow-wave sleep. Although in this case the term could apply to a single trace as well as to many, the inference was (correctly) made that large periodic signals must be caused by underlying neuronal sources acting synchronously. Desynchronized EEG, characteristic of REM sleep and wakefulness, lacks regular periodicity in its global voltage fluctuations.
2. The term “ensemble” has two connotations. First, it can refer to all neurons simultaneously recorded in an experiment; these neurons are not necessarily functionally related or connected. This is the sense meant at this location in the text. Secondly, an ensemble can be understood as a group of neurons exerting a common function. This connotation will be mainly used in chapter 10.
3. In binocular rivalry, two different images are projected onto corresponding areas of the eyes, one image reaching the retina of the left eye and the other image the right eye. Subjects will not report a simultaneous mixing of the two images but instead report a perceptual alternation, sometimes seeing one image in full for a few seconds and the other image at other times. This alternation occurs spontaneously, without any physical change in the stimuli presented.
4. In contrast to anatomic or synaptic connectivity, which relies on morphological evidence for neural connections, functional connectivity denotes the extent to which different areas or cells interact with each other. It is generally quantified by some measure of the correlation between the signals recorded from neural entities and is an indirect measure of connectivity because it is hard to exclude other sources influencing the measured interactions besides the recorded entities themselves.
5. The BOLD contrast is the main signal pursued in fMRI. When neural tissue becomes more electrically active, cells will consume more energy and thus more oxygen. *In situ*, this oxygen is derived from blood circulation, and in response to a bout of neural activity, the local vasculature of the brain generates a complex hemodynamic response containing both components of increased oxygen loss and an increase in fresh, well-oxygenated blood flow. The BOLD signal arises because oxygen-rich and oxygen-poor blood have slightly different magnetic properties. Although fMRI can be done noninvasively, the temporal and spatial resolutions are low as compared to intracranial electrophysiology or live cell imaging.
6. The power of an EEG signal is computed as the square of the EEG amplitude. The amplitude is calculated as the integral average from the peak of the signal to its trough, across a certain predefined time window of sampling. The power spectrum is a graph plotting the power of the EEG in a limited frequency band as a function of frequency, showing the relative strength of each band for the epoch under study.
7. In psychology, a “set” is defined as a pattern of anticipation that predisposes individuals to have enhanced sensitivity to a particular type of input and to respond accordingly. For instance, a “motor set” will facilitate a certain type of response to a stimulus because of a preestablished expectation toward the upcoming stimulus and response.

Chapter 8

1. Iconic memory is a fast-decaying form of sensory short-term memory that is characterized as precategorical and having a high memory capacity. Having been mostly studied in the visual domain, it is distinct from other short-term or working memory by its short duration (< 1 second; Sperling, 1960).

2. Observations made by Grafé, cited in Zeki (1993, p. 318). See also M. von Soden (1932).
3. One may object that “duck” and “rabbit” are learned categories, shaped by many individual examples presented to us over our lifetime. It has become clear, however, that color constancy and judgment of color similarity are also experience-dependent processes. Especially early postnatal development is an indispensable period (Sugita, 2004; Hurlbert, 2007).
4. A similar argument arises from phylogenetic development. Self-recognition as indexed by the mirror mark test is known to occur in relatively few, highly evolved species, such as elephants, monkeys, and bottlenose dolphins (Reiss & Marino, 2001; Plotnik et al., 2006), whereas general perceptual awareness can be argued to be found in many more species (see also chapter 11). Some authors in the literature on self-awareness use a different terminology for consciousness than used in this book. For instance, Lewis (2003) refers to self-awareness or the realization that “I know I know” as (explicit) “consciousness,” whereas general perceptual awareness (without self-reference) is referred to as “knowing” or “implicit consciousness.”
5. Similar out-of-body hallucinations and other dissociative effects have been reported for phencyclidine (PCP, “angel dust”) and the over-the-counter anticoUGH medication dextromethorphan (Rosse et al., 1994; Romanelli & Smith, 2003).
6. Also Damasio (2000) maintains a distinction between consciousness and the protoself, which he regards as a “preconscious biological precedent” of the sense of self. Nonetheless, the representation or “image” of an organism’s own state lies at the root of his concept of consciousness, because consciousness hinges on the production of an imaged, nonverbal account of the effects of an external object on the organism’s state.

Chapter 9

1. The discussion of coexisting reference frames does not fully address the question of whether some sensory maps correlate more to conscious experience than do others, and thus gain more dominance of one’s first-person perspective. Neural activity in V1, as we saw in chapter 7, is important for generating visual experiences, but a range of properties of V1 neurons, including their retinotopic organization and dependence on eye movement, led Rees et al. (2002) and others to conclude that V1 activity does not directly correlate to conscious experience. Higher-order visual areas, such as posterior parietal cortex, would be more plausible candidate regions for determining one’s visual perspective. However, in this scenario it remains to be solved how these areas achieve this dominance, or, in other words, what capacities lower areas such as V1 would be lacking in failing to generate correlates of conscious experience.

If we were to go along with the idea that only high-order, but not primary, cortical maps are conducive to conscious experience, we would also face the puzzle of why representations in cortical areas placed at the same level of the visual hierarchy (e.g., VIP, LIP, medial superior temporal area [MST], and posterior inferotemporal area [PIT]; Felleman & Van Essen, 1991) should give rise to only one coherent worldview, rather than four different perspectives at the same time. One answer may hold that we may be considering a fake problem here: as long as sensory inputs are processed normally and our motor system responds appropriately, there is “nobody” (no agent or separate entity) in the brain “to complain” (cf. Dennett, 1991). Another approach is to explain perceptual coherence from the idea that one reference frame is dominant under normal conditions and suppresses other frames, noting that it may be violated under pathological conditions such as autotomy and out-of-body experience.

2. Whereas this principle of a unique and bidirectional neural-perceptual mapping is probably natural to accept as a ground attitude on representation for many neuroscientists, it is more controversial among philosophers. Representationalist theories, for instance, attribute the brain’s representational power to its relations to external affairs (Dretske, 2003; see also chapter 11). This has the uncomfortable consequence that fully identical brains, generating exactly the same activity patterns, can sustain different conscious experiences.
3. The experiment of lesioning all higher visual and parietal-prefrontal areas has not been done (but see Nakamura & Mishkin, 1986, for a partial approach) and would soon run into the trouble of having no options for behavioral reportability. However, one may instead attempt to evoke a percept of a single, isolated attribute in the absence of other attributes or submodalities being activated. For instance, motion sensations can be evoked by TMS on the scalp overlying area V5 in humans (e.g., Antal et al., 2004). At first sight, an isolated attribute-specific percept seems to favor a scenario in which a feature can be perceived in isolation from other (sub)modalities: motion percepts may not *have* to be integrated with other features into a whole object. According to the current framework, however, the V5 neurons activated by the TMS pulse will send their spikes to other areas and exert influences on these connected neurons, arguing that the activation of the visual

motion-processing pathway was not as isolated as it seemed to be. Neurons in neighboring areas were firing in the background all along, and so the inputs from V5 neurons will have been capable of inducing phase shifts in these baseline patterns. Thus, the relational effects enabling motion perception may well have been preserved but just failed to co-occur with activity bumps coding for other features besides motion.

Chapter 10

1. It should be recalled here that Marr's (1982) original work did not have the explicit goal of elucidating the neural basis of consciousness.
2. We might also refer to the philosophical position of "emergentism" here, were it not that this concept has often been used to refer to collective processes of which the basis is not or is ill understood. In general, an emergent property of a system is understood to be more than the sum of the system's parts, including their individual properties. Consciousness can be claimed to "emerge" from a system of interacting sensory–cognitive modules, but this claim does not add anything new to the more specific elaboration on the functioning of these modules and their interactions. Another problem with emergentism is that it is not unambiguous on the issue of whether emergent properties carry some degree of independence from the basic features of the underlying system. If independence is claimed, holding that emergent properties can exist that cannot be deduced from their basis, the question remains how these properties can arise in the first place. Kim (1993, 1999) has objected to this position by arguments of overdetermination and causal closure. More appropriate than "emergence" may be to label the adjective "conscious" as a supervenient property of a representational system, with dependence on the underlying low-level elements.
3. Technically, the term "ensemble recording" only implies that spike trains from many neurons are recorded at the same time, and that different spikes can be attributed to distinct neurons within the recorded population (Wilson & McNaughton, 1993). In practice, an ensemble recording picks up a limited sample of the entire array of neurons in a brain structure, and the cells in this sample are often highly heterogeneous. In contrast, an "ensemble" as used in figure 10.2 forms a functional entity, a group of neurons exerting a common function.
4. Here the term "multimodal" in the context of such larger meta-networks is to be taken broadly, including not only sensory modalities but also memory and motor systems.
5. Supervenience refers to relationships where high-level properties of a system depend on, and are grounded in, lower-level properties of that system. This position does not imply a strong reductive relationship. For example, global safety of air travel is supervenient on the physical properties of pilots, airplanes, and so forth but cannot be fully captured by physical descriptions because socioeconomic factors also play a major role. In the current context, supervenience holds that not all psychological phenomena can be fully reduced to, or described by, physical phenomena.
6. Ants were chosen in this example as a tribute to Hofstadter's (1979, 1985) ant-colony metaphor. The argument here, however, is different because it critically hinges on the robot's inability to distinguish a veridical ant from a simulating robot agent.

Chapter 11

1. Although C-fibers and pain are a favorite example in discussions on identity theory, it is nowadays recognized as being misguided. The nervous system possesses other types of fiber that are also involved in transmission of nociceptive signals, and C-fibers have been implicated in conveying more than nociceptive information (e.g., warmth, itch, and sensual touch).
2. O'Regan and Noë (2001) acknowledge this problem as they distinguish visual awareness from the qualitative nature of vision per se. In addition to the mastery of visuomotor contingencies, they add that visual awareness also depends on the use of these skills for the purposes of thought, planning, or decision making. However, this addition defers the problem of awareness (or consciousness in the current terminology) to defining what thought and planning are beyond the exercise of sensorimotor contingencies. Either it must be the case that thought and planning are captured under some definition of sensorimotor contingencies (in which case the problem of automaticity persists) or they are different processes. The current account favors the latter view, because, for instance, thought comprises imagery and/or covert linguistic activity one is aware of, and both of these processes by far exceed sensorimotor contingencies in representational complexity.
3. O'Regan and Noë (2001) further defended their viewpoint on dreaming, imagery, and hallucinations arguing that the alternative—that of the brain containing "internal pictorial representations"—does not make sense.

Indeed it is nonsensical to suppose that “to see red, there must be red neurons in the brain” (p. 947). Crucially, however, this is not the only alternative to externalism. The current account posits that the brain *generates* representations by its internal mechanisms, but the result of this operation is that representations are of a *projectional* nature such that they model the world outside the brain, so that neurons do not have to be red in order to render an object red and situate it externally.

4. Further arguments against externalism can be identified, of which the phenomenon of synesthesia presents a strong one. In grapheme synesthesia, for instance, a person will perceive displayed letters or numbers in particular colors, with the letter identity being consistently coupled to the color across time. The identity of the perceived color is not plausibly dictated by a sensorimotor contingency.

5. Vitalism as a philosophical position, prevalent in the late nineteenth and early twentieth centuries, posited that life is made up of something more—a mysterious life-donating force or *élan vital*—than the biochemical processes involving proteins, DNA, RNA, lipid membranes, and so forth (e.g., Bergson, 1911; see also chapter 4).

6. This position is different from a particular “projectivist view” that plays a role in debates on primary versus secondary qualities and the question where qualitative properties should be located (see, e.g., Shoemaker, 1990). One form of projectivism posits that we project, in our experiences, qualitative properties (e.g., colors), or similarities and differences between them, onto things in the world. In chapter 9, however, I argued that experienced information from all sensory modalities is integrated into a situational representation that is of a projectional nature. Shoemaker (1990) is correct in concluding that this would make our experiences systematically illusory, but on the other hand we have no means to discover that a trick is played (unless by abstract inference; see chapter 10). The “trick” can be more positively cast as the brain’s “best guess” about what is going on in the world and our own body.

7. This situation is somewhat comparable to the case of anencephalic patients, discussed by Merker (2007). His thesis that these patients should be considered conscious can be questioned, as already argued in note 2 of chapter 6. In the present context, it is added that his argument is purely based on motor behavior, not on systems requirements for representation.

References

- Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3(Suppl), 1178–1183. doi:10.1038/81453.
- Abeles, M. (1991). *Corticonics, neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.
- Adrian, E. D., & Matthews, B. H. (1934). The interpretation of potential waves in the cortex. *Journal of Physiology*, 81(4), 440–471.
- Aertsen, A. M., Gerstein, G. L., Habib, M. K., & Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of “effective connectivity.” *Journal of Neurophysiology*, 61(5), 900–917.
- Ahissar, E., & Arieli, A. (2001). Figuring space by time. *Neuron*, 32(2), 185–201.
- Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, 10, 25–61.
- Alexander, G. E., Crutcher, M. D., & DeLong, M. R. (1990). Basal ganglia–thalamocortical circuits: parallel substrates for motor, oculomotor, “prefrontal” and “limbic” functions. *Progress in Brain Research*, 85, 119–146.
- Alkire, M. T., Hudetz, A. G., & Tononi, G. (2008). Consciousness and anesthesia. *Science*, 322(5903), 876–880. doi:10.1126/science.1149213.
- Alkire, M. T., McReynolds, J. R., Hahn, E. L., & Trivedi, A. N. (2007). Thalamic microinjection of nicotine reverses sevoflurane-induced loss of righting reflex in the rat. *Anesthesiology*, 107(2), 264–272. doi:10.1097/01.anes.0000270741.33766.24.
- Alkire, M. T., & Miller, J. (2005). General anesthesia and the neural correlates of consciousness. *Progress in Brain Research*, 150, 229–244. doi:10.1016/S0079-6123(05)50017-7.
- Allan, L. G. (1979). The perception of time. *Perception & Psychophysics*, 26, 340–354.
- Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local–global comparisons in visual neurons. *Annual Review of Neuroscience*, 8, 407–430. doi:10.1146/annurev.ne.08.030185.002203.
- Allman, J. M., & Kaas, J. H. (1971). Representation of the visual field in striate and adjoining cortex of the owl monkey (*Aotus trivirgatus*). *Brain Research*, 35(1), 89–106.
- Alter, T., & Howell, R. J. (2012). *Consciousness and the mind-body problem: a reader*. Oxford: Oxford University Press.
- Altmann, C. F., Getzmann, S., & Lewald, J. (2012). Allocentric or craniocentric representation of acoustic space: an electrotomography study using mismatch negativity. *PLoS ONE*, 7(7), e41872. doi:10.1371/journal.pone.0041872.
- Amaral, D. G., & Witter, M. P. (1995). Hippocampal formation. In G. Paxinos (Ed.), *The rat nervous system* (pp. 443–493). London: Academic Press.
- Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo–haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, 4(3), 324–330.
- Amoore, J. E. (1964). Current status of the steric theory of odor. *Annals of the New York Academy of Sciences*, 116, 457–476.

- Andersen, B. B., Gundersen, H. J., & Pakkenberg, B. (2003). Aging of the human cerebellum: a stereological study. *Journal of Comparative Neurology*, 466(3), 356–365. doi:10.1002/cne.10884.
- Andersen, P., Bliss, T. V., Lomo, T., Olsen, L. I., & Skrede, K. K. (1969). Lamellar organization of hippocampal excitatory pathways. *Acta Physiologica Scandinavica*, 76(1), 4A–5A.
- Andersen, R. A., & Cui, H. (2009). Intention, action planning, and decision making in parietal-frontal circuits. *Neuron*, 63(5), 568–583.
- Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230(4724), 456–458.
- Anderson, D. J. (2012). Optogenetics, sex, and violence in the brain: implications for psychiatry. *Biological Psychiatry*, 71(12), 1081–1089. doi:10.1016/j.biopsych.2011.11.012.
- Andreasen, N. C., O'Leary, D. S., Flaum, M., Nopoulos, P., Watkins, G. L., Boles Ponto, L. L., et al. (1997). Hypofrontality in schizophrenia: distributed dysfunctional circuits in neuroleptic-naïve patients. *Lancet*, 349(9067), 1730–1734.
- Anscombe, G. E. M. (1965). The intentionality of sensation: a grammatical feature. In R. J. Butler (Ed.), *Analytical philosophy: second series* (pp. 158–180). Oxford: Basil Blackwell.
- Antal, A., Nitsche, M. A., Kincses, T. Z., Lampe, C., & Paulus, W. (2004). No correlation between moving phosphenes and motor thresholds: a transcranial magnetic stimulation study. *Neuroreport*, 15(2), 297–302.
- Apps, R., & Garwicz, M. (2005). Anatomical and physiological foundations of cerebellar information processing. *Nature Reviews Neuroscience*, 6(4), 297–311. doi:10.1038/nrn1646.
- Aras, N., Oommen, B. J., & Altinel, I. K. (1999). Kohonen network incorporating explicit statistics and its application to the travelling salesman problem. *Neural Networks*, 12(9), 1273–1284.
- Arbib, M. A. (2001). Co-evolution of human consciousness and language. *Annals of the New York Academy of Sciences*, 929, 195–220.
- Arena, P., De Fiore, S., & Patane, L. (2009). Cellular Nonlinear Networks for the emergence of perceptual states: application to robot navigation control. *Neural Networks*, 22(5–6), 801–811. doi:10.1016/j.neunet.2009.06.024.
- Armstrong, D. M. (1980). The Nature of Mind. In N. Block (Ed.), *Readings in philosophy of psychology* (pp. 191–199). Cambridge, MA: Harvard University Press.
- Armstrong, D. (1993). *A materialist theory of the mind*. London: Routledge.
- Armstrong, W. E., Wang, L., Li, C., & Teruyama, R. (2010). Performance, properties and plasticity of identified oxytocin and vasopressin neurones in vitro. *Journal of Neuroendocrinology*, 22(5), 330–342. doi:10.1111/j.1365-2826.2010.01989.x.
- Arzy, S., Thut, G., Mohr, C., Michel, C. M., & Blanke, O. (2006). Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *Journal of Neuroscience*, 26(31), 8074–8081.
- Ascoli, G. A., Alonso-Nanclares, L., Anderson, S. A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., et al. (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience*, 9(7), 557–568. doi:10.1038/nrn2402.
- Aserinsky, E., & Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 118(3062), 273–274.
- Aspect, A. (1999). Bell's inequality test: more ideal than ever. *Nature*, 398, 189–190.
- Attemeave, F. (1961). In defense of homunculi. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 777–782). Cambridge, MA: MIT Press.
- Austin, J. L. (1962). *Sense and sensibilia*. Oxford: Oxford University Press.
- Avendano, C., Rausell, E., Perez-Aguilar, D., & Isorna, S. (1988). Organization of the association cortical afferent connections of area 5: a retrograde tracer study in the cat. *Journal of Comparative Neurology*, 278(1), 1–33. doi:10.1002/cne.902780102.
- Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), 13172–13177. doi:10.1073/pnas.162485599.

- Avillac, M., Deneve, S., Olivier, E., Pouget, A., & Duhamel, J. R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nature Neuroscience*, 8(7), 941–949.
- Axmacher, N., Elger, C. E., & Fell, J. (2008). Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain*, 131(Pt 7), 1806–1817. doi:10.1093/brain/awn103.
- Ayer, A. J. (1953). *The foundations of empirical knowledge*. London: MacMillan.
- Baars, B. J. (1983). Conscious contents provide the nervous system with coherent, global information. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 3, pp. 41–79). New York: Plenum Press.
- Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, 6(1), 47–52.
- Baars, B. J., Ramsay, T. Z., & Laureys, S. (2003). Brain, conscious experience and the observing self. *Trends in Neurosciences*, 26(12), 671–675. doi:10.1016/j.tins.2003.09.015.
- Bachmann, T. (2006). Microgenesis of perception: conceptual, psychophysical and neurobiological aspects. In H. Ögmen & B. G. Breitmeyer (Eds.), *The first half second—the microgenesis and temporal dynamics of unconscious and conscious visual processes* (pp. 11–33). Cambridge, MA: MIT Press.
- Baird, A. D., Wilson, S. J., Bladin, P. F., Saling, M. M., & Reutens, D. C. (2007). Neurological control of human sexual behaviour: insights from lesion studies. *Journal of Neurology, Neurosurgery, and Psychiatry*, 78(10), 1042–1049. doi:10.1136/jnnp.2006.107193.
- Baker, S. N., Olivier, E., & Lemon, R. N. (1997). Coherent oscillations in monkey motor cortex and hand muscle EMG show task-dependent modulation. *Journal of Physiology*, 501(Pt 1), 225–241.
- Ballard, D. H., Hinton, G. E., & Sejnowski, T. J. (1983). Parallel visual computation. *Nature*, 306(5938), 21–26.
- Bannister, A. P. (2005). Inter- and intra-laminar connections of pyramidal cells in the neocortex. *Neuroscience Research*, 53(2), 95–103. doi:10.1016/j.neures.2005.06.019.
- Banno, T., Ichinohe, N., Rockland, K. S., & Komatsu, H. (2011). Reciprocal connectivity of identified color-processing modules in the monkey inferior temporal cortex. *Cerebral Cortex*, 21(6), 1295–1310. doi:10.1093/cercor/bhq211.
- Barbour, D. L., & Callaway, E. M. (2008). Excitatory local connections of superficial neurons in rat auditory cortex. *Journal of Neuroscience*, 28(44), 11174–11185. doi:10.1523/JNEUROSCI.2093-08.2008.
- Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1(4), 371–394.
- Barnes, R. D. (1980). *Invertebrate zoology* (4th ed.). Philadelphia: Saunders College.
- Barnett, L., Buckley, C. L., & Bullock, S. (2011). Neural complexity: a graph theoretic interpretation. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 83(4 Pt 1), 041906.
- Barrett, A. B., Murphy, M., Bruno, M. A., Noirhomme, Q., Boly, M., Laureys, S., et al. (2012). Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. *PLoS ONE*, 7(1), e29072. doi:10.1371/journal.pone.0029072.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–609; discussion 610–660.
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695–711. doi:10.1016/j.neuron.2012.10.038.
- Battaglia, F. P., Benchenane, K., Sirota, A., Pennartz, C. M., & Wiener, S. I. (2011). The hippocampus: hub of brain network communication for memory. *Trends in Cognitive Sciences*, 15(7), 310–318. doi:10.1016/j.tics.2011.05.008.
- Battaglia, F. P., & Pennartz, C. M. (2011). The construction of semantic memory: grammar-based representations learned from relational episodic information. *Frontiers in Computational Neuroscience*, 5, 36. doi:10.3389/fncom.2011.00036.
- Battaglia, F. P., Sutherland, G. R., & McNaughton, B. L. (2004). Hippocampal sharp wave bursts coincide with neocortical “up-state” transitions. *Learning & Memory*, 11(6), 697–704. doi:10.1101/lm.73504.
- Bauby, J. D. (1998). *The diving bell and the butterfly: a memoir of life in death*. New York: Vintage Books, Random House.

- Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nature Reviews Neuroscience*, 3(7), 563–573.
- Baynes, K., & Gazzaniga, M. S. (2000). Consciousness, introspection and the split brain: the two minds/one body problem. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1355–1363). Cambridge, MA: MIT Press.
- Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience*, 18(1), 428–437.
- Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.
- Bell, A. H., Meredith, M. A., Van Opstal, A. J., & Munoz, D. P. (2005). Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *Journal of Neurophysiology*, 93(6), 3659–3673.
- Bell, C. (1869). Idea of a new anatomy of the brain; Submitted for the observations of his friends (1811); reprinted in *Journal of Anatomy and Physiology*, 3, 154–157.
- Benito-Gutiérrez, E. (2006). A gene catalogue of the amphioxus nervous system. *International Journal of Biological Sciences*, 2(3), 149–160.
- Berger, H. (1929). Über das Elektroenzephalogramm des Menschen (On the electroencephalogram of man). *Archiv für Psychiatrie und Nervenkrankheiten*, 87, 527–570.
- Bergson, H. (1911). *Matter and memory*. New York: Macmillan.
- Berkeley, G. (1707). *An essay towards a new theory of vision*. Dublin: Pepyat.
- Berridge, K. C., & Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology*, 199(3), 457–480. doi:10.1007/s00213-008-1099-6.
- Bevan, M. D., Magill, P. J., Terman, D., Bolam, J. P., & Wilson, C. J. (2002). Move to the rhythm: oscillations in the subthalamic nucleus–external globus pallidus network. *Trends in Neurosciences*, 25(10), 525–531.
- Bhatia, K. P., & Marsden, C. D. (1994). The behavioural and motor consequences of focal lesions of the basal ganglia in man. *Brain*, 117(Pt 4), 859–876.
- Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2(1), 32–48.
- Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J., & Ralph, M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex*, 20(11), 2728–2738. doi:10.1093/cercor/bhq019.
- Binzegger, T., Douglas, R. J., & Martin, K. A. (2004). A quantitative map of the circuit of cat primary visual cortex. *Journal of Neuroscience*, 24(39), 8441–8453. doi:10.1523/JNEUROSCI.1400-04.2004.
- Birbach, O., Frese, U., & Bauml, B. (2011). *Realtime perception for catching a flying ball with a mobile humanoid*. Paper presented at the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China.
- Bischof, M., & Bassetti, C. L. (2004). Total dream loss: a distinct neuropsychological dysfunction after bilateral PCA stroke. *Annals of Neurology*, 56(4), 583–586. doi:10.1002/ana.20246.
- Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Stimulating illusory own-body perceptions. *Nature*, 419(6904), 269–270.
- Blasdel, G. G. (1992). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, 12(8), 3139–3161.
- Blethyn, K. L., Hughes, S. W., Tóth, T. I., Cope, D. W., & Crunelli, V. (2006). Neuronal basis of the slow (<1 Hz) oscillation in neurons of the nucleus reticularis thalami in vitro. *Journal of Neuroscience*, 26(9), 2474–2486.
- Bliss, T. V., & Gardner-Medwin, A. R. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, 232(2), 357–374.

- Bliss, T. V., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, 232(2), 331–356.
- Block, N. (1980a). *Readings in philosophy of psychology*. Cambridge, MA: Harvard University Press.
- Block, N. (1980b). Troubles with functionalism. In N. Block (Ed.), *Readings in philosophy of psychology* (Vol. 1, pp. 268–305). Cambridge, MA: Harvard University Press.
- Block, N. (1990). Consciousness and accessibility. *Behavioral and Brain Sciences*, 13, 596–598.
- Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, 9(2), 46–52.
- Blonder, L. X., Heilman, K. M., Ketterson, T., Rosenbek, J., Raymer, A., Crosson, B., et al. (2005). Affective facial and lexical expression in a prosodic versus aphasic stroke patients. *Journal of the International Neuropsychological Society*, 11(6), 677–685. doi:10.1017/S1355617705050794.
- Blouet, C., & Schwartz, G. J. (2010). Hypothalamic nutrient sensing in the control of energy homeostasis. *Behavioural Brain Research*, 209(1), 1–12.
- Boeijinga, P. H., & Lopes da Silva, F. H. (1988). Differential distribution of beta and theta EEG activity in the entorhinal cortex of the cat. *Brain Research*, 448(2), 272–286.
- Bolanowski, S. J., Jr., & Doty, R. W. (1987). Perceptual “blankout” of monocular homogeneous fields (Ganzfelder) is prevented with binocular viewing. *Vision Research*, 27(6), 967–982.
- Bonelli, R. M., & Cummings, J. L. (2007). Frontal–subcortical circuitry and behavior. *Dialogues in Clinical Neuroscience*, 9(2), 141–151.
- Boraud, T., Bezard, E., Bioulac, B., & Gross, C. E. (2002). From single extracellular unit recording in experimental and human Parkinsonism to the development of a functional concept of the role played by the basal ganglia in motor control. *Progress in Neurobiology*, 66(4), 265–283.
- Boring, E. G. (1933). *The physical dimensions of consciousness*. New York: Appleton-Century.
- Boring, E. G. (1950). *A history of experimental psychology* (2nd ed.). New York: Appleton Century Crofts.
- Bosman, C. A., Lansink, C. S., & Pennartz, C. M. (2014). Functions of gamma-band synchronization in cognition: from single circuits to functional diversity across cortical and subcortical systems. *European Journal of Neuroscience*, 39(11), 1982–1999. doi:10.1111/ejn.12606.
- Bosman, C. A., Schoffelen, J. M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., et al. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron*, 75(5), 875–888. doi:10.1016/j.neuron.2012.06.037.
- Bourque, C. W. (2008). Central mechanisms of osmosensation and systemic osmoregulation. *Nature Reviews Neuroscience*, 9(7), 519–531.
- Bouyer, J. J., Montaron, M. F., Buser, P., Durand, C., & Rougeul, A. (1992). Effects of mediodorsalis thalamic nucleus lesions on vigilance and attentive behaviour in cats. *Behavioural Brain Research*, 51(1), 51–60.
- Braitenberg, V., & Schüz, A. (1991). *Anatomy of the cortex: statistics and geometry*. New York: Springer Verlag.
- Brancucci, A., Franciotti, R., D'Anselmo, A., Della Penna, S., & Tommasi, L. (2011). The sound of consciousness: neural underpinnings of auditory perception. *Journal of Neuroscience*, 31(46), 16611–16618. doi:10.1523/JNEUROSCI.3949-11.2011.
- Braun, A. R., Balkin, T. J., Wesensten, N. J., Carson, R. E., Varga, M., Baldwin, P., et al. (1997). Regional cerebral blood flow throughout the sleep–wake cycle: an H₂(15)O PET study. *Brain*, 120(Pt 7), 1173–1197.
- Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: detection and discrimination tasks. *Perception & Psychophysics*, 60(1), 1–23.
- Breen, N., Caine, D., & Coltheart, M. (2000). Models of face recognition and delusional misidentification: a critical review. *Cognitive Neuropsychology*, 17(1), 55–71. doi:10.1080/026432900380481.
- Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K., et al. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*, 29(1), 287–296.
- Bressler, S. L., & Kelso, J. A. (2001). Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, 5(1), 26–36.

- Broersen, L. M., Heinsbroek, R. P., de Bruin, J. P., Uylings, H. B., & Olivier, B. (1995). The role of the medial prefrontal cortex of rats in short-term memory functioning: further support for involvement of cholinergic, rather than dopaminergic mechanisms. *Brain Research*, 674(2), 221–229.
- Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, 2(1), 51–61. doi:10.1038/35049064.
- Brown, T. H., Wong, R. K., & Prince, D. A. (1979). Spontaneous miniature synaptic potentials in hippocampal neurons. *Brain Research*, 177(1), 194–199.
- Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46(2), 369–384.
- Brugge, P. (2002). Reflective mirrors: perspective-taking in autoscopic phenomena. *Cognitive Neuropsychiatry*, 7(3), 179–194. doi:10.1080/13546800244000076.
- Brugge, P., Kolllias, S. S., Muri, R. M., Crelier, G., Hepp-Reymond, M. C., & Regard, M. (2000). Beyond remembering: phantom sensations of congenitally absent limbs. *Proceedings of the National Academy of Sciences of the United States of America*, 97(11), 6167–6172. doi:10.1073/pnas.100510697.
- Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, 64(2), 123–152.
- Bryson, A., & Ho, Y.-C. (1969). *Applied optimal control*. New York: Blaisdell.
- Buck, L., & Axel, R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, 65(1), 175–187.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124, 1–38. doi:10.1196/annals.1440.011.
- Bullier, J. (2001a). Feedback connections and conscious vision. *Trends in Cognitive Sciences*, 5(9), 369–370.
- Bullier, J. (2001b). Integrated model of visual processing. *Brain Research. Brain Research Reviews*, 36(2–3), 96–107.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4, 73–121.
- Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315(5820), 1860–1862. doi:10.1126/science.1138071.
- Buzsaki, G. (1989). Two-stage model of memory trace formation: a role for “noisy” brain states. *Neuroscience*, 31(3), 551–570.
- Buzsaki, G. (2006). *Rhythms of the brain*. Oxford: Oxford University Press.
- Buzsaki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68(3), 362–385. doi:10.1016/j.neuron.2010.09.023.
- Buzsaki, G., & Chrobak, J. J. (2005). Synaptic plasticity and self-organization in the hippocampus. *Nature Neuroscience*, 8(11), 1418–1420. doi:10.1038/nn1105-1418.
- Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926–1929. doi:10.1126/science.1099745.
- Buzsaki, G., Horvath, Z., Urioste, R., Hetke, J., & Wise, K. (1992). High-frequency network oscillation in the hippocampus. *Science*, 256(5059), 1025–1027.
- Calis, G. (1984). Concerning Gibson's "On the face of it": immediate perception and single-glance face recognition. *Acta Psychologica*, 55, 195–214.
- Calis, G. J. J., Sterenborg, J. M., & Maarse, F. J. (1984). Initial microgenetic steps in single-glance face recognition. *Acta Psychologica*, 55, 215–230.
- Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex*, 11(12), 1110–1123.
- Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio–visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, 14(2), 427–438. doi:10.1006/nimg.2001.0812.

- Campolattaro, M. M., Kashev, A., Lee, I., & Freeman, J. H. (2011). Neuronal correlates of cross-modal transfer in the cerebellum and pontine nuclei. *Journal of Neuroscience*, 31(11), 4051–4062. doi:10.1523/JNEUROSCI.4142-10.2011.
- Camprodon, J. A., Zohary, E., Brodbeck, V., & Pascual-Leone, A. (2010). Two phases of V1 activity for visual recognition of natural images. *Journal of Cognitive Neuroscience*, 22(6), 1262–1269. doi:10.1162/jocn.2009.21253.
- Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*, 22(11), 2886–2902.
- Cardin, J. A., Carlen, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., et al. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature*, 459(7247), 663–667. doi:10.1038/nature08002.
- Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, 26(3), 321–352.
- Carmichael, S. T., Clugnet, M. C., & Price, J. L. (1994). Central olfactory connections in the macaque monkey. *Journal of Comparative Neurology*, 346(3), 403–434. doi:10.1002/cne.903460306.
- Carruthers, P. (1989). Brute experience. *Journal of Philosophy*, 86(5), 258–269.
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine*, 5(198), 198ra105. doi:10.1126/scitranslmed.3006294.
- Castellucci, V. F., Carew, T. J., & Kandel, E. R. (1978). Cellular analysis of long-term habituation of the gill-withdrawal reflex of *Aplysia californica*. *Science*, 202(4374), 1306–1308.
- Chafee, M. V., & Goldman-Rakic, P. S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *Journal of Neurophysiology*, 83(3), 1550–1566.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2, 200–219.
- Chalmers, D. J. (1996). *The conscious mind*. Oxford: Oxford University Press.
- Chalmers, D. J. (2004). The representational character of experience. In B. Leiter (Ed.), *The future for philosophy* (pp. 153–181). Oxford: Oxford University Press.
- Cherubini, E., Herrling, P. L., Lanfumey, L., & Stanzione, P. (1988). Excitatory amino acids in synaptic excitation of rat striatal neurones in vitro. *Journal of Physiology*, 400, 677–690.
- Cherubini, E., Gaiarsa, J. L., & Ben-Ari, Y. (1991). GABA: an excitatory transmitter in early postnatal life. *Trends in Neurosciences*, 14(12), 515–519.
- Chevalier, G., & Deniau, J. M. (1990). Disinhibition as a basic process in the expression of striatal functions. *Trends in Neurosciences*, 13(7), 277–280.
- Cheyne, D., Gaetz, W., Garner, L., Lachaux, J. P., Ducorps, A., Schwartz, D., et al. (2003). Neuromagnetic imaging of cortical oscillations accompanying tactile stimulation. *Brain Research. Cognitive Brain Research*, 17(3), 599–611.
- Chiba, T., Kayahara, T., & Nakano, K. (2001). Efferent projections of infralimbic and prelimbic areas of the medial prefrontal cortex in the Japanese monkey, *Macaca fuscata*. *Brain Research*, 888(1), 83–101.
- Ching, S., Purdon, P. L., Vijayan, S., Kopell, N. J., & Brown, E. N. (2012). A neurophysiological-metabolic model for burst suppression. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), 3095–3100. doi:10.1073/pnas.1121461109.
- Cho, J., & West, M. O. (1997). Distributions of single neurons related to body parts in the lateral striatum of the rat. *Brain Research*, 756(1–2), 241–246.
- Choi, G. B., Stettler, D. D., Kallman, B. R., Bhaskar, S. T., Fleischmann, A., & Axel, R. (2011). Driving opposing behaviors with ensembles of piriform neurons. *Cell*, 146(6), 1004–1015. doi:10.1016/j.cell.2011.07.041.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

- Chow, B. Y., Han, X., Dobry, A. S., Qian, X., Chuong, A. S., Li, M., et al. (2010). High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature*, 463(7277), 98–102. doi:10.1038/nature08652.
- Christakou, A., Robbins, T. W., & Everitt, B. J. (2005). Prolonged neglect following unilateral disruption of a prefrontal cortical-dorsal striatal system. *European Journal of Neuroscience*, 21(3), 782–792.
- Churchland, P. M. (1984). *Matter and consciousness*. Cambridge, MA: MIT Press.
- Churchland, P. M. (1995). *The engine of reason, the seat of the soul*. Cambridge, MA: MIT Press.
- Churchland, P. S. (2002). Self-representation in nervous systems. *Science*, 296(5566), 308–310. doi:10.1126/science.1070564.
- Churchland, P. S. (1986). *Neurophilosophy: toward a unified science of the mind-brain*. Cambridge, MA: MIT Press.
- Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.
- Clapham, J. C. (2012). Central control of thermogenesis. *Neuropharmacology*, 63(1), 111–123. doi:10.1016/j.neuropharm.2011.10.014.
- Clark, A. (2000). *A theory of sentience*. Oxford: Oxford University Press.
- Coesmans, M., Weber, J. T., De Zeeuw, C. I., & Hansel, C. (2004). Bidirectional parallel fiber plasticity in the cerebellum under climbing fiber control. *Neuron*, 44(4), 691–700. doi:10.1016/j.neuron.2004.10.031.
- Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, 15(8), 358–364. doi:10.1016/j.tics.2011.06.008.
- Cohen, M. R., & Newsome, W. T. (2004). What electrical microstimulation has revealed about the neural basis of cognition. *Current Opinions in Neurobiology*, 14(2), 169–177.
- Cohen, Y. E., Cohen, I. S., & Gifford, G. W. (2004). Modulation of LIP activity by predictive auditory and visual cues. *Cerebral Cortex*, 14(12), 1287–1301.
- Colgin, L. L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., et al. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*, 462(7271), 353–357. doi:10.1038/nature08573.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240–247.
- Colton, B. P. (1903). *Zoology: descriptive and practical*. Boston: D.C. Heath.
- Comella, C. L., & Shannon, K. M. (1999). Hemiballismus. *Current Treatment Options in Neurology*, 1(1), 1–5.
- Committeri, G., Galati, G., Paradis, A. L., Pizzamiglio, L., Berthoz, A., & LeBihan, D. (2004). Reference frames for spatial cognition: different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location. *Journal of Cognitive Neuroscience*, 16(9), 1517–1535.
- Conway, B. R., & Tsao, D. Y. (2006). Color architecture in alert macaque cortex revealed by fMRI. *Cerebral Cortex*, 16(11), 1604–1613. doi:10.1093/cercor/bhj099.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, 3(3), 201–215. doi:10.1038/nrn755.
- Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nature Reviews. Neuroscience*, 3(2), 153–160. doi:10.1038/nrn726.
- Cossart, R., Aronov, D., & Yuste, R. (2003). Attractor dynamics of network UP states in the neocortex. *Nature*, 423(6937), 283–288. doi:10.1038/nature01614.
- Cowey, A., & Heywood, C. A. (1997). Cerebral achromatopsia: colour blindness despite wavelength processing. *Trends in Cognitive Sciences*, 1(4), 133–139. doi:10.1016/S1364-6613(97)01043-7.
- Cowey, A., & Walsh, V. (2000). Magnetically induced phosphene in sighted, blind and blindsighted observers. *Neuroreport*, 11(14), 3269–3273.
- Craft, E., Schutze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure–ground organization. *Journal of Neurophysiology*, 97(6), 4310–4326.
- Craig, A. D. (2002). How do you feel? Interception: the sense of the physiological condition of the body. *Nature Reviews. Neuroscience*, 3(8), 655–666.

- Crick, F., & Koch, C. (1990). Toward a neurobiological theory of consciousness. *Seminars in Neuroscience*, 2, 263–275.
- Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6(2), 119–126.
- Crick, F. C., & Koch, C. (2005). What is the function of the claustrum? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1458), 1271–1279. doi:10.1098/rstb.2005.1661.
- Critchley, H. D., & Rolls, E. T. (1996). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *Journal of Neurophysiology*, 75(4), 1673–1686.
- Croner, L. J., & Albright, T. D. (1999). Seeing the big picture: integration of image cues in the primate visual system. *Neuron*, 24(4), 777–789.
- Crunelli, V., & Hughes, S. W. (2010). The slow (<1 Hz) rhythm of non-REM sleep: a dialogue between three cardinal oscillators. *Nature Neuroscience*, 13(1), 9–17.
- Cruse, D., Chennu, S., Chatelle, C., Bekinschtein, T. A., Fernandez-Espejo, D., Pickard, J. D., et al. (2012). Bedside detection of awareness in the vegetative state: a cohort study. *Lancet*, 378(9809), 2088–2094. doi:10.1016/S0140-6736(11)61224-5.
- Curran, H. V., & Monaghan, L. (2001). In and out of the K-hole: a comparison of the acute and residual effects of ketamine in frequent and infrequent ketamine users. *Addiction (Abingdon, England)*, 96(5), 749–760. doi:10.1080/0962140020039116.
- D'Angelo, E., & De Zeeuw, C. I. (2009). Timing and plasticity in the cerebellum: focus on the granular layer. *Trends in Neurosciences*, 32(1), 30–40. doi:10.1016/j.tins.2008.09.007.
- Dalley, J. W., Cardinal, R. N., & Robbins, T. W. (2004). Prefrontal executive and cognitive functions in rodents: neural and neurochemical substrates. *Neuroscience and Biobehavioral Reviews*, 28(7), 771–784.
- Dalley, J. W., Theobald, D. E., Bouger, P., Chudasama, Y., Cardinal, R. N., & Robbins, T. W. (2004). Cortical cholinergic function and deficits in visual attentional performance in rats following 192 IgG-saporin-induced lesions of the medial prefrontal cortex. *Cerebral Cortex*, 14(8), 922–932. doi:10.1093/cercor/bhh052.
- Dalton, P., Doolittle, N., Nagata, H., & Breslin, P. A. (2000). The merging of the senses: integration of subthreshold taste and smell. *Nature Neuroscience*, 3(5), 431–432. doi:10.1038/74797.
- Damasio, A. (2000). *The feeling of what happens*. London: Vintage.
- Darian-Smith, I. (Ed.). (1984). *Handbook of physiology: a critical, comprehensive presentation of physiological knowledge and concepts. Section 1, the nervous system, sensory processes* (Vol. 3). Bethesda, MD: American Physiological Society.
- Darian-Smith, I., Sugitani, M., Heywood, J., Karita, K., & Goodwin, A. (1982). Touching textured surfaces: cells in somatosensory cortex respond both to finger movement and to surface features. *Science*, 218(4575), 906–909.
- Daselaar, S. M., Huijbers, W., de Jonge, M., Goltstein, P. M., & Pennartz, C. M. (2010). Experience-dependent alterations in conscious resting state activity following perceptuomotor learning. *Neurobiology of Learning and Memory*, 93(3), 422–427. doi:10.1016/j.nlm.2009.12.009.
- Daselaar, S. M., Porat, Y., Huijbers, W., & Pennartz, C. M. (2010). Modality-specific and modality-independent components of the human imagery system. *NeuroImage*, 52(2), 677–685. doi:10.1016/j.neuroimage.2010.04.239.
- Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron*, 63(4), 497–507. doi:10.1016/j.neuron.2009.07.027.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8, 1704–1711.
- Dayan, P. (2006). Images, frames, and connectionist hierarchies. *Neural Computation*, 18(10), 2293–2319.
- De Jeu, M., & Pennartz, C. (2002). Circadian modulation of GABA function in the rat suprachiasmatic nucleus: excitatory effects during the night phase. *Journal of Neurophysiology*, 87(2), 834–844.
- de Lafuente, V., & Romo, R. (2006). Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proceedings of the National Academy of Sciences of the United States of America*, 103(39), 14266–14271. doi:10.1073/pnas.0605826103.

- De Sanctis, P., Ritter, W., Molholm, S., Kelly, S. P., & Foxe, J. J. (2008). Auditory scene analysis: the interaction of stimulation rate and frequency separation on pre-attentive grouping. *European Journal of Neuroscience*, 27(5), 1271–1276. doi:10.1111/j.1460-9568.2008.06080.x.
- De Zeeuw, C. I., Hoebeek, F. E., Bosman, L. W., Schonewille, M., Witter, L., & Koekkoek, S. K. (2011). Spatiotemporal firing patterns in the cerebellum. *Nature Reviews. Neuroscience*, 12(6), 327–344.
- Dean, P., Porrill, J., Ekerot, C. F., & Jorntell, H. (2010). The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. *Nature Reviews. Neuroscience*, 11(1), 30–43. doi:10.1038/nrn2756.
- Dear, S. P., Simmons, J. A., & Fritz, J. (1993). A possible neuronal basis for representation of acoustic scenes in auditory cortex of the big brown bat. *Nature*, 364(6438), 620–623. doi:10.1038/364620a0.
- Decavel, C., & Van den Pol, A. N. (1990). GABA: a dominant neurotransmitter in the hypothalamus. *Journal of Comparative Neurology*, 302(4), 1019–1037. doi:10.1002/cne.903020423.
- Deco, G., & Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 44(6), 621–642.
- Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227. doi:10.1016/j.neuron.2011.03.018.
- Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, 10(5), 204–211. doi:10.1016/j.tics.2006.03.007.
- Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*, 95(24), 14529–14534.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79(1–2), 1–37.
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4(7), 752–758. doi:10.1038/89551.
- Dehay, C., Kennedy, H., & Bullier, J. (1988). Characterization of transient cortical projections from auditory, somatosensory, and motor cortices to visual areas 17, 18, and 19 in the kitten. *Journal of Comparative Neurology*, 272(1), 68–89.
- Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, 5(10), e260. doi:10.1371/journal.pbio.0050260.
- Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, 2(8), 740–745. doi:10.1038/11205.
- Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4(8), 826–831. doi:10.1038/90541.
- Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology, Paris*, 98(1–3), 249–258. doi:10.1016/j.jphysparis.2004.03.011.
- Dennett, D. C. (1983). Artificial intelligence and the strategies of psychological investigation. In J. Miller (Ed.), *States of mind* (pp. 66–81). London: British Broadcasting Corporation.
- Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown.
- Dennett, D. C. (2003). The self as a responding-and responsible-artifact. *Annals of the New York Academy of Sciences*, 1001, 39–50.
- Dennett, D. C. (2005). *Sweet dreams: philosophical obstacles to a science of consciousness*. Cambridge, MA: MIT Press.
- Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer. *Behavioral and Brain Sciences*, 15(2), 183–247.
- Derdikman, D., & Moser, E. I. (2010). A manifold of spatial maps in the brain. *Trends in Cognitive Sciences*, 14(12), 561–569.
- Descartes, R. (1641). Meditations. Reprinted in *Discourse on method and the meditations* (translated by F. E. Sutcliffe, 1986). Harmondsworth, UK: Penguin.

- Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, 4(8), 2051–2062.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222. doi:10.1146/annurev.ne.18.030195.001205.
- Destexhe, A., Hughes, S. W., Rudolph, M., & Crunelli, V. (2007). Are corticothalamic “up” states fragments of wakefulness? *Trends in Neurosciences*, 30(7), 334–342.
- Devinsky, O. (2009). Delusional misidentifications and duplications: right brain lesions, left brain delusions. *Neurology*, 72(1), 80–87. doi:10.1212/01.wnl.0000338625.47892.74.
- Devinsky, O., Morrell, M. J., & Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain*, 118(Pt 1), 279–306.
- DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, 11(5), 219–226.
- Di Prisco, G. V., & Freeman, W. J. (1985). Odor-related bulbar EEG spatial pattern analysis during appetitive conditioning in rabbits. *Behavioral Neuroscience*, 99(5), 964–978.
- Diba, K., & Buzsaki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, 10(10), 1241–1242. doi:10.1038/nrn1961.
- Dickman, M. S. (2001). von Economo encephalitis. *Archives of Neurology*, 58(10), 1696–1698.
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2), 114–126. doi:10.1038/nrn2762.
- Dierks, T., Linden, D. E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., et al. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Neuron*, 22(3), 615–621.
- Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, 33(4), 1331–1336a. doi:10.1523/JNEUROSCI.4081-12.2013.
- Douglas, R. J., & Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, 27, 419–451.
- Douglas, R. J., & Martin, K. A. (2007). Mapping the matrix: the ways of neocortex. *Neuron*, 56(2), 226–238. doi:10.1016/j.neuron.2007.10.017.
- Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Dretske, F. (2003). Experience as representation. *Philosophical Issues*, 13, 67–82.
- Dreyfus, H. L. (1979). *What computers can't do*. Cambridge, MA: MIT Press.
- Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition*, 80(1–2), 61–95.
- Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience*, 1(1), 17–22. doi:10.1038/217.
- Driver, J., & Vuilleumier, P. (2001). Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition*, 79(1–2), 39–88.
- Duistermars, B. J., & Frye, M. A. (2010). Multisensory integration for odor tracking by flying *Drosophila*: behavior, circuits and speculation. *Communicative & Integrative Biology*, 3(1), 60–63.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General*, 113(4), 501–517.
- Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, 3(Suppl), 1184–1191. doi:10.1038/81460.
- Eby, S. A., Buchner, E. J., Bryant, M. G., & Mak, H. K. (2012). The rehabilitation of Anton syndrome. *PM & R*, 4(5), 385–387. doi:10.1016/j.pmrj.2011.12.012.
- Eccles, J. C., Ito, M., & Szentágothai, J. (1967). *The cerebellum as a neuronal machine*. Berlin: Springer.
- Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., & Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965), 584–587. doi:10.1126/science.1179867.

- Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., et al. (1988). Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biological Cybernetics*, 60(2), 121–130.
- Edelman, G. M., Gally, J. A., & Baars, B. J. (2011). Biology of consciousness. *Frontiers in Psychology*, 2, 4. doi:10.3389/fpsyg.2011.00004.
- Edwards, R., Xiao, D., Keysers, C., Foldiak, P., & Perrett, D. (2003). Color sensitivity of cells responsive to complex stimuli in the temporal cortex. *Journal of Neurophysiology*, 90(2), 1245–1256. doi:10.1152/jn.00524.2002.
- Egan, G., Silk, T., Zamarripa, F., Williams, J., Federico, P., Cunningham, R., et al. (2003). Neural correlates of the emergence of consciousness of thirst. *Proceedings of the National Academy of Sciences of the United States of America*, 100(25), 15241–15246. doi:10.1073/pnas.2136650100.
- Eggermont, J. J. (1998). Is there a neural code? *Neuroscience and Biobehavioral Reviews*, 22(2), 355–370.
- Eggermont, J. J. (2001). Between sound and perception: reviewing the search for a neural code. *Hearing Research*, 157(1–2), 1–42.
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nature Reviews. Neuroscience*, 1(1), 41–50. doi:10.1038/35036213.
- Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., & Lipton, P. (2011). Towards a functional organization of episodic memory in the medial temporal lobe. *Neuroscience and Biobehavioral Reviews*. doi:10.1016/j.neubiorev.2011.07.006.
- Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, 30, 123–152. doi:10.1146/annurev.neuro.30.051606.094328.
- Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., et al. (2003). Cellular networks underlying human spatial navigation. *Nature*, 425(6954), 184–188.
- Ellis, H. D., & Lewis, M. B. (2001). Capgras delusion: a window on face recognition. *Trends in Cognitive Sciences*, 5(4), 149–156.
- Engel, A. K., & Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Current Opinion in Neurobiology*, 20(2), 156–165. doi:10.1016/j.conb.2010.02.015.
- Engel, A. K., Konig, P., & Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding. *Proceedings of the National Academy of Sciences of the United States of America*, 88(20), 9136–9140.
- Ergenoglu, T., Demiralp, T., Bayraktaroglu, Z., Ergen, M., Beydagl, H., & Uresin, Y. (2004). Alpha rhythm of the EEG modulates visual detection performance in humans. *Brain Research. Cognitive Brain Research*, 20(3), 376–383. doi:10.1016/j.cogbrainres.2004.03.009.
- Ermentrout, G. B., & Cowan, J. D. (1979). A mathematical theory of visual hallucination patterns. *Biological Cybernetics*, 34(3), 137–150.
- Eskenasy, A. C., & Clarke, S. (2000). Hierarchy within human SI: supporting data from cytochrome oxidase, acetylcholinesterase and NADPH-diaphorase staining patterns. *Somatosensory & Motor Research*, 17(2), 123–132.
- Eslinger, P. J., & Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology*, 35(12), 1731–1741.
- Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, 318(5853), 1147–1150. doi:10.1126/science.1148979.
- Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, 22(13), 5749–5759.
- Farah, M. (2004). *Visual agnosia* (2nd ed.). Cambridge, MA: MIT Press.
- Feldmeyer, D., & Sakmann, B. (2000). Synaptic efficacy and reliability of excitatory connections between the principal neurones of the input (layer 4) and output layer (layer 5) of the neocortex. *Journal of Physiology*, 525(Pt 1), 31–39.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), 1–47.

- Fernandez-Espejo, D., & Owen, A. M. (2013). Detecting awareness after severe brain injury. *Nature Reviews Neuroscience*, 14(11), 801–809. doi:10.1038/nrn3608.
- Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B. A., Angelini, G., et al. (2010). Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 107(6), 2681–2686. doi:10.1073/pnas.0913008107.
- Ferrucci, D. (2010). *Build Watson: an overview of DeepQA for the Jeopardy! challenge*. Paper presented at the 19th international conference on Parallel architectures and compilation techniques, Vienna, Austria.
- Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. T. (2013). Watson: beyond Jeopardy! *Artificial Intelligence*, 199, 93–105. doi:10.1016/j.artint.2012.06.009.
- Feyerabend, P. (1963). Mental events and the brain. *Journal of Philosophy*, 60, 295–296.
- ffytche, D. H., & Zeki, S. (2011). The primary visual cortex, and feedback to it, are not necessary for conscious vision. *Brain*, 134(Pt 1), 247–257. doi:10.1093/brain/awq305.
- Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., et al. (2009). Neural “ignition”: enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron*, 64(4), 562–574. doi:10.1016/j.neuron.2009.11.001.
- Fitzgerald, P. J., Lane, J. W., Thakur, P. H., & Hsiao, S. S. (2006). Receptive field (RF) properties of the macaque second somatosensory cortex: rf size, shape, and somatotopic organization. *Journal of Neuroscience*, 26(24), 6473–6484.
- Flanagan-Cato, L. M. (2011). Sex differences in the neural circuit that mediates female sexual receptivity. *Frontiers in Neuroendocrinology*, 32(2), 124–136.
- Fodor, J. (1975). *The language of thought*. Hassocks: Harvester Press.
- Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3, 63–109.
- Fodor, J. (1983). *Modularity of mind: an essay on faculty psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1985). Précis of the modularity of mind. *Behavioral and Brain Sciences*, 8, 1–42.
- Fodor, J. (1987). *Psychosemantics: the problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Ford, J. M., Roach, B. J., Jorgensen, K. W., Turner, J. A., Brown, G. G., Notestine, R., et al. (2009). Tuning in to the voices: a multisite fMRI study of auditory hallucinations. *Schizophrenia Bulletin*, 35(1), 58–66. doi:10.1093/schbul/sbn140.
- Forman, S. A., & Miller, K. W. (2011). Anesthetic sites and allosteric mechanisms of action on Cys-loop ligand-gated ion channels. *Canadian Journal of Anaesthesia*, 58(2), 191–205. doi:10.1007/s12630-010-9419-9.
- Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084), 680–683. doi:10.1038/nature04587.
- Foulkes, W. D. (1962). Dream reports from different stages of sleep. *Journal of Abnormal and Social Psychology*, 65, 14–25.
- Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, 6(2), 119–130. doi:10.1038/nrn1607.
- Franks, K. M., & Isaacson, J. S. (2006). Strong single-fiber sensory inputs to olfactory cortex: implications for olfactory coding. *Neuron*, 49(3), 357–363. doi:10.1016/j.neuron.2005.12.026.
- Franks, N. P. (2008). General anaesthesia: from molecular targets to neuronal pathways of sleep and arousal. *Nature Reviews Neuroscience*, 9(5), 370–386. doi:10.1038/nrn2372.
- Fraser, A., & Wilcox, K. J. (1979). Perception of illusory movement. *Nature*, 281(5732), 565–566.
- Freedman, D. J., & Miller, E. K. (2008). Neural mechanisms of visual categorization: insights from neurophysiology. *Neuroscience and Biobehavioral Reviews*, 32(2), 311–329. doi:10.1016/j.neubiorev.2007.07.011.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, 23(12), 5235–5246.

- Freeman, W. J. (1996). Random activity at the microscopic neural level in cortex ("noise") sustains and is regulated by low-dimensional dynamics of macroscopic cortical activity ("chaos"). *International Journal of Neural Systems*, 7(4), 473–480.
- Freeman, W. J. (2007). Indirect biological measures of consciousness from field studies of brains as dynamical systems. *Neural Networks*, 20(9), 1021–1031. doi:10.1016/j.neunet.2007.09.004.
- Freeman, W. M. D., & Watts, J. W. M. D. (1937). Prefrontal lobotomy in the treatment of mental disorders. *Southern Medical Journal*, 30, 23–31.
- Freud, S. (1900). *Die Traumdeutung (The interpretation of dreams)*. Leipzig: F. Deuticke.
- Friedman, D. P. (1983). Laminar patterns of termination of cortico-cortical afferents in the somatosensory system. *Brain Research*, 273(1), 147–151.
- Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science*, 269(5225), 853–855.
- Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291(5508), 1560–1563. doi:10.1126/science.291.5508.1560.
- Frijda, N. (2006). *The laws of emotion*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fu, K. M., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., et al. (2003). Auditory cortical neurons respond to somatosensory stimulation. *Journal of Neuroscience*, 23(20), 7510–7515.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, 7, 773–778.
- Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2), 331–349.
- Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1993). Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas." *Journal of Neuroscience*, 13(4), 1479–1497.
- Funnell, M. G., Corballis, P. M., & Gazzaniga, M. S. (2000). Cortical and subcortical interhemispheric interactions following partial and complete callosotomy. *Archives of Neurology*, 57(2), 185–189.
- Fuster, J. M. (1997). Network memory. *Trends in Neurosciences*, 20(10), 451–459.
- Fuster, J. M. (2001). The prefrontal cortex—an update: time is of the essence. *Neuron*, 30(2), 319–333.
- Fuster, J. M. (2003). *Cortex and mind—unifying cognition*. Oxford: Oxford University Press.
- Gabor, A. J., & Peele, T. L. (1964). Alterations of behavior following stimulation of the claustrum of the cat. *Electroencephalography and Clinical Neurophysiology*, 17, 513–519.
- Galati, G., Lobel, E., Vallar, G., Berthoz, A., Pizzamiglio, L., & Le Bihan, D. (2000). The neural basis of egocentric and allocentric coding of space in humans: a functional magnetic resonance study. *Experimental Brain Research*, 133(2), 156–164.
- Galati, G., Pelle, G., Berthoz, A., & Committeri, G. (2010). Multiple reference frames used by the human brain for spatial perception and memory. *Experimental Brain Research*, 206(2), 109–120. doi:10.1007/s00221-010-2168-8.
- Galletti, C., Battaglini, P. P., & Fattori, P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Experimental Brain Research*, 96(2), 221–229.
- Garcia-Rill, E. (1991). The pedunculopontine nucleus. *Progress in Neurobiology*, 36(5), 363–389. doi:0301–0082(91)90016-T [pii].
- Gaudry, Q., & Kristan, W. B., Jr. (2012). Decision points: the factors influencing the decision to feed in the medicinal leech. *Frontiers in Neuroscience*, 6, 101. doi:10.3389/fnins.2012.00101.
- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, 12(3), 495–504.
- Gazzaniga, M. S., Holtzman, J. D., & Smylie, C. S. (1987). Speech without conscious awareness. *Neurology*, 37(4), 682–685.

- Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *Journal of Neuroscience*, 21(23), 9430–9437. doi:10.1523/JNEUROSCI.2000-01.
- Gentet, L. J. (2012). Functional diversity of supragranular GABAergic neurons in the barrel cortex. *Frontiers in Neural Circuits*, 6, 52. doi:10.3389/fncir.2012.00052.
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6), 278–285.
- Gho, M., & Varela, F. J. (1988). A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *Journal de Physiologie*, 83(2), 95–101.
- Giannakopoulos, P., Gold, G., Duc, M., Michel, J. P., Hof, P. R., & Bouras, C. (1999). Neuroanatomic correlates of visual agnosia in Alzheimer's disease: a clinicopathologic study. *Neurology*, 52(1), 71–77.
- Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Erlbaum.
- Gilbert, C. D. (1983). Microcircuitry of the visual cortex. *Annual Review of Neuroscience*, 6, 217–247. doi:10.1146/annurev.ne.06.030183.001245.
- Girardeau, G., & Zugardo, M. (2011). Hippocampal ripples and memory consolidation. *Current Opinion in Neurobiology*, 21(3), 452–459. doi:10.1016/j.conb.2011.02.005.
- Gläscher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., et al. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 109(36), 14681–14686. doi:10.1073/pnas.1206608109.
- Gold, J. J., & Squire, L. R. (2006). The anatomy of amnesia: neurohistological analysis of three new cases. *Learning & Memory*, 13(6), 699–710. doi:10.1101/lm.357406.
- Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, 14(3), 477–485.
- Goll, J. C., Crutch, S. J., & Warren, J. D. (2010). Central auditory disorders: toward a neuropsychology of auditory objects. *Current Opinion in Neurology*, 23(6), 617–627. doi:10.1097/WCO.0b013e32834027f6.
- Goltstein, P. M., Coffey, E. B., Roelfsema, P. R., & Pennartz, C. M. (2013). In vivo two-photon Ca²⁺ imaging reveals selective reward effects on stimulus-specific assemblies in mouse visual cortex. *Journal of Neuroscience*, 33(28), 11540–11555. doi:10.1523/JNEUROSCI.1341-12.2013.
- Goltstein, P. M., Montijn, J. S., & Pennartz, C. M. A. (2011). Spatial organization and temporal dynamics of visual cortex assemblies during the awake and anesthetized state. *Society for Neuroscience Abstracts Program* 730.02.
- Goodale, M. A., Kroliczak, G., & Westwood, D. A. (2005). Dual routes to action: contributions of the dorsal and ventral streams to adaptive behavior. *Progress in Brain Research*, 149, 269–283.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.
- Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, 349(6305), 154–156. doi:10.1038/349154a0.
- Gottfried, J. A. (2010). Central mechanisms of odour object perception. *Nature Reviews. Neuroscience*, 11(9), 628–641.
- Grace, A. A., & Bunney, B. S. (1984). The control of firing pattern in nigral dopamine neurons: burst firing. *Journal of Neuroscience*, 4(11), 2877–2890.
- Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*, 17(2), 484–495. doi:10.1016/j.concog.2008.03.019.
- Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: still alive and well. *Neuron*, 24(1), 31–47.
- Gray, C. M., Konig, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213), 334–337. doi:10.1038/338334a0.
- Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 86(5), 1698–1702.

- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition*, 125(1), 125–130. doi:10.1016/j.cognition.2012.06.007.
- Graziano, M. S. (1999). Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Sciences of the United States of America*, 96(18), 10418–10421.
- Graziano, M. S., Reiss, L. A., & Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. *Nature*, 397(6718), 428–430.
- Greenberg, D. S., Houweling, A. R., & Kerr, J. N. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nature Neuroscience*, 11(7), 749–751. doi:10.1038/nn.2140.
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 290(1038), 181–197.
- Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), 253–258.
- Grillner, S., Hellgren, J., Ménard, A., Saitoh, K., & Wikström, M. A. (2005). Mechanisms for selection of basic motor programs: roles of the striatum and pallidum. *Trends in Neurosciences*, 28(7), 364–370.
- Groenewegen, H. J., & Berendse, H. W. (1994). The specificity of the “nonspecific” midline and intralaminar thalamic nuclei. *Trends in Neurosciences*, 17(2), 52–57.
- Groenewegen, H. J., & Uylings, H. B. (2000). The prefrontal cortex and the integration of sensory, limbic and autonomic information. *Progress in Brain Research*, 126, 3–28. doi:10.1016/S0079-6123(00)26003-2.
- Groos, G., & Mason, R. (1978). Maintained discharge of rat suprachiasmatic neurons at different adaptation levels. *Neuroscience Letters*, 8(1), 59–64. doi:0304–3940(78)90098–8 [pii].
- Grosmark, A. D., Mizuseki, K., Pastalkova, E., Diba, K., & Buzsaki, G. (2012). REM sleep reorganizes hippocampal excitability. *Neuron*, 75(6), 1001–1007. doi:10.1016/j.neuron.2012.08.015.
- Gross, C. G. (2002). Genealogy of the “grandmother cell.” *Neuroscientist*, 8(5), 512–518.
- Gross, C. G., & Sergent, J. (1992). Face recognition. *Current Opinion in Neurobiology*, 2(2), 156–161.
- Grossberg, S. (2007). Consciousness CLEARS the mind. *Neural Networks*, 20(9), 1040–1053. doi:10.1016/j.neunet.2007.09.014.
- Grossberg, S. (2013). Adaptive resonance theory: how a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, 37, 1–47. doi:10.1016/j.neunet.2012.09.017.
- Groves, P. M. (1983). A theory of the functional organization of the neostriatum and the neostriatal control of voluntary movement. *Brain Research*, 286(2), 109–132.
- Guenther, F. H., Brumberg, J. S., Wright, E. J., Nieto-Castanon, A., Tourville, J. A., Panko, M., et al. (2009). A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE*, 4(12).
- Guillery, R. W., & Sherman, S. M. (2002). Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system. *Neuron*, 33(2), 163–175.
- Gupta, A. S., van der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, 65(5), 695–705. doi:10.1016/j.neuron.2010.01.034.
- Haass, C., & Selkoe, D. J. (2007). Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nature Reviews. Molecular Cell Biology*, 8(2), 101–112. doi:10.1038/nrm2101.
- Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience*, 20(6), 2369–2382.
- Haber, S. N., Lynd, E., Klein, C., & Groenewegen, H. J. (1990). Topographic organization of the ventral striatal efferent projections in the rhesus monkey: an anterograde tracing study. *Journal of Comparative Neurology*, 293(2), 282–298.
- Haberly, L. B. (2001). Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry. *Chemical Senses*, 26(5), 551–576.
- Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hearing Research*, 271(1–2), 133–146. doi:10.1016/j.heares.2010.01.011.

- Haider, B., Häusser, M., & Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature*, 493(7430), 97–100. doi:10.1038/nature11665.
- Haith, A., & Vijayakumar, S. (2009). Implications of different classes of sensorimotor disturbance for cerebellar-based motor learning models. *Biological Cybernetics*, 100(1), 81–95. doi:10.1007/s00422-008-0266-5.
- Haken, H. (2006). Synergetics of brain function. *International Journal of Psychophysiology*, 60(2), 110–124. doi:10.1016/j.ijpsycho.2005.12.006.
- Halligan, P. W., Hunt, M., Marshall, J. C., & Wade, D. T. (1995). Sensory detection without localization. *Neurocase*, 1, 259–266.
- Hampson, R. E., Song, D., Opris, I., Santos, L. M., Shin, D. C., Gerhardt, G. A., et al. (2013). Facilitation of memory encoding in primate hippocampus by a neuroprosthesis that promotes task-specific neural firing. *Journal of Neural Engineering*, 10(6), 066013. doi:10.1088/1741-2560/10/6/066013.
- Hanley, J. R., Young, A. W., & Pearson, N. A. (1991). Impairment of the visuo-spatial sketch pad. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 43(1), 101–125.
- Hari, R., & Salmelin, R. (1997). Human cortical oscillations: a neuromagnetic view through the skull. *Trends in Neurosciences*, 20(1), 44–49. doi:10.1016/S0166-2236(96)10065-5.
- Harley, C. M., Cienfuegos, J., & Wagenaar, D. A. (2011). Developmentally regulated multisensory integration for prey localization in the medicinal leech. *Journal of Experimental Biology*, 214(Pt 22), 3801–3807. doi:10.1242/jeb.059618.
- Harnad, S. (1990). The symbol grounding problem. *Physica D. Nonlinear Phenomena*, 42, 335–346.
- Harrington, D. L., Haaland, K. Y., & Knight, R. T. (1998). Cortical networks underlying mechanisms of time perception. *Journal of Neuroscience*, 18(3), 1085–1095.
- Harris, K. D. (2005). Neural signatures of cell assembly organization. *Nature Reviews. Neuroscience*, 6(5), 399–407. doi:10.1038/nrn1669.
- Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., & Buzsaki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, 424(6948), 552–556. doi:10.1038/nature01834.
- Harthoorn, L. F., Sane, A., Nethe, M., & Van Heerikhuize, J. J. (2005). Multi-transcriptional profiling of melanin-concentrating hormone and orexin-containing neurons. *Cellular and Molecular Neurobiology*, 25(8), 1209–1223. doi:10.1007/s10571-005-8184-8.
- Häusser, M., Raman, I. M., Otis, T., Smith, S. L., Nelson, A., du Lac, S., et al. (2004). The beat goes on: spontaneous firing in mammalian neuronal microcircuits. *Journal of Neuroscience*, 24(42), 9215–9219. doi:10.1523/JNEUROSCI.3375-04.2004.
- Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. New York: Wiley.
- Helmholtz, H. von (1863). *Die Lehre von den Tonempfindungen*. Braunschweig: Vieweg.
- Helmholtz, H. von (1866). *Handbuch der physiologischen Optik*. Leipzig: Leopold Voss.
- Hering, E. (1874). *Outlines of a theory of the light sense*. Cambridge, MA: Harvard University Press.
- Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.
- Heywood, C. A., Cowey, A., & Newcombe, F. (1991). Chromatic discrimination in a cortically colour blind observer. *European Journal of Neuroscience*, 3(8), 802–812.
- Heywood, C. A., Nicholas, J. J., & Cowey, A. (1996). Behavioural and electrophysiological chromatic and achromatic contrast sensitivity in an achromatopsic patient. *Journal of Neurology, Neurosurgery, and Psychiatry*, 60(6), 638–643.
- Higuchi, T., & Okere, C. O. (2002). Role of the supraoptic nucleus in regulation of parturition and milk ejection revisited. *Microscopy Research and Technique*, 56(2), 113–121. doi:10.1002/jemt.10016.
- Hilgetag, C. C., O'Neill, M. A., & Young, M. P. (1996). Indeterminate organization of the visual system. *Science*, 271(5250), 776–777.
- Hintikka, K. J. J. (1969). On the logic of perception. In N. S. Care & R. H. Grimm (Eds.), *Perception and personal identity* (pp. 140–173). Cleveland, OH: Case Western Reserve University Press.

- Hobson, J. A. (2005). Sleep is of the brain, by the brain and for the brain. *Nature*, 437(7063), 1254–1256. doi:10.1038/nature04283.
- Hobson, J. A. (2009). REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews Neuroscience*, 10(11), 803–813. doi:10.1038/nrn2716.
- Hobson, J. A., & Pace-Schott, E. F. (2002). The cognitive neuroscience of sleep: neuronal systems, consciousness and learning. *Nature Reviews Neuroscience*, 3(9), 679–693. doi:10.1038/nrn915.
- Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398), 372–375. doi:10.1038/Nature11076.
- Hodges, J. R., & Patterson, K. (2007). Semantic dementia: a unique clinicopathological syndrome. *Lancet Neurology*, 6(11), 1004–1014. doi:10.1016/S1474-4422(07)70266-1.
- Hofbauer, R. K., Rainville, P., Duncan, G. H., & Bushnell, M. C. (2001). Cortical representation of the sensory dimension of pain. *Journal of Neurophysiology*, 86(1), 402–411.
- Hoffman, K. L., & McNaughton, B. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science*, 297(5589), 2070–2073.
- Hofstadter, D. R. (1979). *Godel, Escher, Bach, an eternal golden braid*. New York: Basic Books.
- Hofstadter, D. R. (1985). *Metamagical themes: questing for the essence of mind and pattern*. New York: Basic Books.
- Hollerman, J. R., Tremblay, L., & Schultz, W. (2000). Involvement of basal ganglia and orbitofrontal cortex in goal-directed behavior. *Progress in Brain Research*, 126, 193–215. doi:10.1016/S0079-6123(00)26015-9.
- Holmes, G. (1939). The cerebellum of man. *Brain*, 62, 1–30.
- Holtmaat, A., Wilbrecht, L., Knott, G. W., Welker, E., & Svoboda, K. (2006). Experience-dependent and cell-type-specific spine growth in the neocortex. *Nature*, 441(7096), 979–983.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8), 2554–2558.
- Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 81(10), 3088–3092.
- Hopfield, J. J. (1994). An envisioning of consciousness. *Science*, 263(5147), 696. doi:10.1126/science.263.5147.696.
- Horel, J. A., Keating, E. G., & Misantone, L. J. (1975). Partial Kluver–Bucy syndrome produced by destroying temporal neocortex or amygdala. *Brain Research*, 94(2), 347–359.
- Horgan, T. (2000). Narrow content and the phenomenology of intentionality. In Presidential Address to the Society for Philosophy and Psychology (Ed.). New York.
- Horn, B. K. P. (1975). Obtaining shape from shading information. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 115–155). New York: McGraw-Hill.
- Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., et al. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain*, 126, 1691–1712.
- Horton, J. C., & Hoyt, W. F. (1991). The representation of the visual field in human striate cortex: a revision of the classic Holmes map. *Archives of Ophthalmology*, 109(6), 816–824.
- Hoshino, O., Zheng, M., & Kuroiwa, K. (2003). Roles of dynamic linkage of stable attractors across cortical networks in recalling long-term memory. *Biological Cybernetics*, 88(3), 163–176. doi:10.1007/s00422-002-0366-6.
- Hsiao, S. (2008). Central mechanisms of tactile shape perception. *Current Opinion in Neurobiology*, 18(4), 418–424. doi:10.1016/j.conb.2008.09.001.
- Hu, A., Zhang, W., & Wang, Z. (2010). Functional feedback from mushroom bodies to antennal lobes in the *Drosophila* olfactory pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 107(22), 10262–10267. doi:10.1073/pnas.0914912107.
- Huang, C. L., & Winer, J. A. (2000). Auditory thalamocortical projections in the cat: laminar and areal patterns of input. *Journal of Comparative Neurology*, 427(2), 302–331.

- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148, 574–591.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106–154.
- Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 198(1130), 1–59.
- Huber, D., Petreanu, L., Ghitani, N., Ranade, S., Hromadka, T., Mainen, Z., et al. (2008). Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. *Nature*, 451(7174), 61–64. doi:10.1038/nature06445.
- Huffman, K. J., & Krubitzer, L. (2001). Area 3a: topographic organization and cortical connections in marmoset monkeys. *Cerebral Cortex*, 11(9), 849–867.
- Hughes, D. P., Andersen, S. B., Hywel-Jones, N. L., Himaman, W., Billen, J., & Boomsma, J. J. (2011). Behavioral mechanisms and morphological symptoms of zombie ants dying from fungal infection. *BMC Ecology*, 11, 13. doi:10.1186/1472-6785-11-13.
- Huijbers, W., Pennartz, C. M., Rubin, D. C., & Daselaar, S. M. (2011). Imagery and retrieval of auditory and visual information: neural correlates of successful and unsuccessful performance. *Neuropsychologia*, 49(7), 1730–1740. doi:10.1016/j.neuropsychologia.2011.02.051.
- Huijbers, W., Vannini, P., Sperling, R. A., Pennartz, C. M. A., Cabeza, R., & Daselaar, S. M. (2012). Explaining the encoding/retrieval flip: memory-related deactivations and activations in the posteromedial cortex. *Neuropsychologia*, 50(14), 3764–3774. doi:10.1016/j.neuropsychologia.2012.08.021.
- Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience*, 25(45), 10420–10436. doi:10.1523/JNEUROSCI.4684-04.2005.
- Hume, D. (1739). *A treatise of human nature*. London: Penguin Books.
- Humphrey, G. K., Goodale, M. A., Corbett, M., & Aglioti, S. (1995). The McCollough effect reveals orientation discrimination in a case of cortical blindness. *Current Biology*, 5(5), 545–551.
- Humphreys, G. W., & Riddoch, M. J. (1987). *To see but not to see: a case study of visual agnosia*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hurlbert, A. (2007). Colour constancy. *Current Biology*, 17(21), R906–R907. doi:10.1016/j.cub.2007.08.022.
- Husserl, E. (1928). *Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins*. Halle, Germany: Max Niemeyer Verlag.
- Huxter, J., Burgess, N., & O'Keefe, J. (2003). Independent rate and temporal coding in hippocampal pyramidal cells. *Nature*, 425(6960), 828–832. doi:10.1038/nature02058.
- Imig, T. J., & Reale, R. A. (1980). Patterns of cortico-cortical connections related to tonotopic maps in cat auditory cortex. *Journal of Comparative Neurology*, 192(2), 293–332. doi:10.1002/cne.901920208.
- Inouye, S. T., & Kawamura, H. (1979). Persistence of circadian rhythmicity in a mammalian hypothalamic "island" containing the suprachiasmatic nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, 76(11), 5962–5966.
- Isaacson, J. S. (2010). Odor representations in mammalian cortical circuits. *Current Opinion in Neurobiology*, 20(3), 328–331. doi:10.1016/j.conb.2010.02.004.
- Ito, R., Robbins, T. W., Pennartz, C. M., & Everitt, B. J. (2008). Functional interaction between the hippocampus and nucleus accumbens shell is necessary for the acquisition of appetitive spatial context conditioning. *Journal of Neuroscience*, 28(27), 6950–6959. doi:10.1523/JNEUROSCI.1615-08.2008.
- Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., et al. (2012). Sound-driven synaptic inhibition in primary visual cortex. *Neuron*, 73(4), 814–828. doi:10.1016/j.neuron.2011.12.026.
- Izawa, J., Criscimagna-Hemminger, S. E., & Shadmehr, R. (2012). Cerebellar contributions to reach adaptation and learning sensory consequences of action. *Journal of Neuroscience*, 32(12), 4230–4239. doi:10.1523/JNEUROSCI.6353-11.2012.
- Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Computation*, 18(2), 245–282. doi:10.1162/089976606775093882.

- Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, 3(7), 272–279.
- Jackendoff, R. (2007). *Language, consciousness, culture: essays on mental structure*. Cambridge, MA: MIT Press.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(127), 127–136. doi:10.2307/2960077.
- JadHAV, S. P., Kemere, C., German, P. W., & Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science*, 336(6087), 1454–1458. doi:10.1126/science.1217230.
- James, T. W., Kim, S., & Fisher, J. S. (2007). The neural basis of haptic object processing. *Canadian Journal of Experimental Psychology*, 61(3), 219–229.
- James, W. (1890). *Principles of psychology*. London: MacMillan.
- Jastrow, J. (1900). *Fact and fable in psychology*. Boston: Houghton Mifflin.
- Jaynes, J. (1990). *The origin of consciousness in the breakdown of the bicameral mind* (2nd ed.). Boston, MA: Houghton-Mifflin.
- Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in Neurosciences*, 18(7), 314–320.
- Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, 10(1), 100–107. doi:10.1038/nn1825.
- Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, 27(45), 12176–12189. doi:10.1523/JNEUROSCI.3761-07.2007.
- Jones, E. G. (2002). Thalamic circuitry and thalamocortical synchrony. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 357(1428), 1659–1673. doi:10.1098/rstb.2002.1168.
- Jones, E. G., Coulter, J. D., & Hendry, S. H. (1978). Intracortical connectivity of architectonic fields in the somatic sensory, motor and parietal cortex of monkeys. *Journal of Comparative Neurology*, 181(2), 291–347.
- Jouvet, M. (1973). Essai sur le rêve (Essay on dreaming). *Archives Italiennes de Biologie*, 111(3–4), 564–576.
- Julesz, B. (1971). *Foundations of Cyclopean perception*. Cambridge, MA: MIT Press.
- Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802), 91–97.
- Kaisti, K. K., Langsjo, J. W., Aalto, S., Oikonen, V., Sipila, H., Teras, M., et al. (2003). Effects of sevoflurane, propofol, and adjunct nitrous oxide on regional cerebral blood flow, oxygen consumption, and blood volume in humans. *Anesthesiology*, 99(3), 603–613.
- Kalenscher, T., Lansink, C. S., Lankelma, J. V., & Pennartz, C. M. (2010). Reward-associated gamma oscillations in ventral striatum are regionally differentiated and modulate local firing activity. *Journal of Neurophysiology*, 103(3), 1658–1672. doi:10.1152/jn.00432.2009.
- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (Eds.). (1991). *Principles of neural science* (3rd ed.). New York: McGraw-Hill.
- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science* (4th ed.). New York: McGraw-Hill.
- Kant, I. (1787). *Kritik der Reinen Vernunft (Critique of pure reason)*. Hamburg: Felix Meiner Verlag.
- Karnath, H. O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, 411(6840), 950–953. doi:10.1038/35082075.
- Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, 365(6443), 250–252. doi:10.1038/365250a0.
- Kasanetz, F., Riquelme, L. A., & Murer, M. G. (2002). Disruption of the two-state membrane potential of striatal neurones during cortical desynchronisation in anaesthetised rats. *Journal of Physiology*, 543(Pt 2), 577–589.
- Kauer, J. S. (2002). On the scents of smell in the salamander. *Nature*, 417(6886), 336–342. doi:10.1038/417336a.
- Kay, L. M., & Freeman, W. J. (1998). Bidirectional processing in the olfactory–limbic axis during olfactory behavior. *Behavioral Neuroscience*, 112(3), 541–553.

- Kayser, C., Montemurro, M. A., Logothetis, N. K., & Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron*, 61(4), 597–608. doi:10.1016/j.neuron.2009.01.008.
- Kelly, R. M., & Strick, P. L. (2004). Macro-architecture of basal ganglia loops with the cerebral cortex: use of rabies virus to reveal multisynaptic circuits. *Progress in Brain Research*, 143, 449–459.
- Kemp, J. M., & Powell, T. P. (1970). The cortico-striate projection in the monkey. *Brain*, 93(3), 525–546.
- Kennedy, H., & Bullier, J. (1985). A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. *Journal of Neuroscience*, 5(10), 2815–2830.
- Kenshalo, D. R., Iwata, K., Sholas, M., & Thomas, D. A. (2000). Response properties and organization of nociceptive neurons in area 1 of monkey primary somatosensory cortex. *Journal of Neurophysiology*, 84(2), 719–729.
- Kesey, K. (1962). *One flew over the cuckoo's nest*. New York: Viking Press.
- Keyser, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature Reviews Neuroscience*, 11(6), 417–428. doi:10.1038/nrn2833.
- Kilgour, A. R., de Gelder, B., & Lederman, S. J. (2004). Haptic face recognition and prosopagnosia. *Neuropsychologia*, 42(6), 707–712. doi:10.1016/j.neuropsychologia.2003.11.021.
- Killcross, S., Robbins, T. W., & Everitt, B. J. (1997). Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. *Nature*, 388(6640), 377–380. doi:10.1038/41097.
- Kim, J. (1993). *Supervenience and mind: selected philosophical essays*. Cambridge: Cambridge University Press.
- Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95, 3–36.
- Kim, J. E., Song, H., Jeong, J. H., Choi, K. G., & Na, D. L. (2007). Bilateral ageusia in a patient with a left ventroposteromedial thalamic infarct: cortical localization of taste sensation by statistical parametric mapping analysis of PET images. *Journal of Clinical Neurology*, 3(3), 161–164. doi:10.3988/jcn.2007.3.3.161.
- King, A. J. (2009). Visual influences on auditory spatial learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1515), 331–339. doi:10.1098/rstb.2008.0230.
- King, A. J., & Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neuroscience*, 12(6), 698–701. doi:10.1038/nn.2308.
- King, C., Henze, D. A., Leinekugel, X., & Buzsaki, G. (1999). Hebbian modification of a hippocampal population pattern in the rat. *Journal of Physiology*, 521(Pt 1), 159–167.
- King, J. R., Sitt, J. D., Faugeras, F., Rohaut, B., El Karoui, I., Cohen, L., et al. (2013). Information sharing in the brain indexes consciousness in noncommunicative patients. *Current Biology*, 23(19), 1914–1919. doi:10.1016/j.cub.2013.07.075.
- Kinney, H. C., Korein, J., Panigrahy, A., Dikkes, P., & Goode, R. (1994). Neuropathological findings in the brain of Karen Ann Quinlan. The role of the thalamus in the persistent vegetative state. *New England Journal of Medicine*, 330(21), 1469–1475.
- Kirouac, G. J., & Pittman, Q. J. (1999). Identification of barosensitive neurons in the mediobasal forebrain using juxtaglomerular labeling. *American Journal of Physiology*, 276(6 Pt 2), R1766–R1771.
- Kita, T., Kita, H., & Kitai, S. T. (1984). Passive electrical membrane properties of rat neostriatal neurons in an in vitro slice preparation. *Brain Research*, 300(1), 129–139. doi:0006-8993(84)91347-7 [pii].
- Kitcher, P. (1988). Marr's computational theory of vision. *Philosophy of Science*, 55, 1–24.
- Klier, E. M., & Angelaki, D. E. (2008). Spatial updating and the maintenance of visual constancy. *Neuroscience*, 156(4), 801–818. doi:10.1016/j.neuroscience.2008.07.079.
- Klimesch, W. (1996). Memory processes, brain oscillations and EEG synchronization. *International Journal of Psychophysiology*, 24(1–2), 61–100.
- Knight, R. T., & Graboweczyk, G. M. (1995). Escape from linear time: prefrontal cortex and conscious experience. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 1319–1339). Cambridge, MA: MIT Press.

- Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjostrom, P. J., & Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345), 87–91. doi:10.1038/nature09880.
- Koch, C. (2004). *The quest for consciousness*. Englewood, CO: Roberts.
- Koch, C. (2012). Consciousness: Confessions of a romantic reductionist. Cambridge, MA: MIT Press.
- Koch, C., & Tononi, G. (2008). Can machines be conscious? *IEEE Spectrum*, 45(6), 54–59.
- Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, 11(1), 16–22. doi:10.1016/j.tics.2006.10.012.
- Koch, G., Oliveri, M., Carlesimo, G. A., & Caltagirone, C. (2002). Selective deficit of time perception in a patient with right prefrontal cortex lesion. *Neurology*, 59(10), 1658–1659.
- Köhler, W. (1940). *Dynamics in psychology*. New York: Grove Press.
- Kohonen, T. (1984). *Self-organization and associative memory*. New York: Springer Verlag.
- König, P., Engel, A. K., Roelfsema, P. R., & Singer, W. (1995). How precise is neuronal synchronization? *Neural Computation*, 7(3), 469–485.
- Koolhaas, J. M. (1978). Hypothalamically induced intraspecific aggressive behaviour in the rat. *Experimental Brain Research*, 32(3), 365–375.
- Kornblith, S., Cheng, X., Ohayon, S., & Tsao, D. Y. (2013). A network for scene processing in the macaque temporal lobe. *Neuron*, 79(4), 766–781. doi:10.1016/j.neuron.2013.06.015.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, 2(9), 635–642. doi:10.1038/35090055.
- Koubeissi, M. Z., Bartolomei, F., Beltagy, A., & Picard, F. (2014). Electrical stimulation of a small brain area reversibly disrupts consciousness. *Epilepsy & Behavior*, 37C, 32–35. doi:10.1016/j.yebeh.2014.05.027.
- Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences*, 14(7), 301–307. doi:10.1016/j.tics.2010.04.006.
- Kouider, S., Sackur, J., & Gardelle, V. (2012). Do we still need phenomenal consciousness? Comment on Block. *Trends in Cognitive Sciences*, 16(3), 140–141, author reply 141–142. doi:10.1016/j.tics.2012.01.003.
- Kreiman, G., Fried, I., & Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 8378–8383. doi:10.1073/pnas.072194099.
- Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, 3(9), 946–953. doi:10.1038/78868.
- Kripke, S. A. (1972). Naming and necessity. In G. Harman & D. Davidson (Eds.), *The semantics of natural language* (pp. 253–355). Dordrecht: Reidel.
- Kudrimoti, H. S., Barnes, C. A., & McNaughton, B. L. (1999). Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *Journal of Neuroscience*, 19(10), 4090–4101.
- Kühn, A. A., Kempf, F., Brucke, C., Gaynor Doyle, L., Martinez-Torres, I., Pogosyan, A., et al. (2008). High-frequency stimulation of the subthalamic nucleus suppresses oscillatory beta activity in patients with Parkinson's disease in parallel with improvement in motor performance. *Journal of Neuroscience*, 28(24), 6165–6173. doi:10.1523/JNEUROSCI.0282-08.2008.
- Kuhn, B., Denk, W., & Bruno, R. M. (2008). In vivo two-photon voltage-sensitive dye imaging reveals top-down control of cortical layers 1 and 2 during wakefulness. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21), 7588–7593. doi:10.1073/pnas.0802462105.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Kuo, H., & Chang, H. T. (1992). Ventral pallido–striatal pathway in the rat brain: a light and electron microscopic study. *Journal of Comparative Neurology*, 321(4), 626–636. doi:10.1002/cne.903210409.
- Kurzweil, R. (2005). *The singularity is near: when humans transcend biology*. New York: Viking.
- Lam, T. K. (2010). Neuronal regulation of homeostasis by nutrient sensing. *Nature Medicine*, 16(4), 392–395. doi:10.1038/nm0410-392.

- Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10(11), 494–501. doi:10.1016/j.tics.2006.09.001.
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571–579.
- Lamme, V. A., & Spekreijse, H. (1998). Neuronal synchrony does not represent texture segregation. *Nature*, 396(6709), 362–366.
- Lamme, V. A., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, 8(4), 529–535.
- Land, E. H., Hubel, D. H., Livingstone, M. S., Perry, S. H., & Burns, M. M. (1983). Colour-generating interactions across the corpus callosum. *Nature*, 303(5918), 616–618.
- Lang, P. J. (1994). The varieties of emotional experience: a meditation on James–Lange theory. *Psychological Review*, 101(2), 211–221.
- Lang, P. J., & Davis, M. (2006). Emotion, motivation, and the brain: reflex foundations in animal and human research. *Progress in Brain Research*, 156, 3–29. doi:10.1016/S0079-6123(06)56001-7.
- Langsjo, J. W., Maksimow, A., Salmi, E., Kaisti, K., Aalto, S., Oikonen, V., et al. (2005). S-ketamine anesthesia increases cerebral blood flow in excess of the metabolic needs in humans. *Anesthesiology*, 103(2), 258–268.
- Lansink, C. S., Goltstein, P. M., Lankelma, J. V., Joosten, R. N., McNaughton, B. L., & Pennartz, C. M. (2008). Preferential reactivation of motivationally relevant information in the ventral striatum. *Journal of Neuroscience*, 28(25), 6372–6382. doi:10.1523/JNEUROSCI.1054-08.2008.
- Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L., & Pennartz, C. M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biology*, 7(8), e1000173. doi:10.1371/journal.pbio.1000173.
- Lansink, C. S., Jackson, J. C., Lankelma, J. V., Ito, R., Robbins, T. W., Everitt, B. J., et al. (2012). Reward cues in space: commonalities and differences in neural coding by hippocampal and ventral striatal ensembles. *Journal of Neuroscience*, 32(36), 12444–12459. doi:10.1523/JNEUROSCI.0593-12.2012.
- Laplane, D., Baulac, M., Widlocher, D., & Dubois, B. (1984). Pure psychic akinesia with bilateral lesions of basal ganglia. *Journal of Neurology, Neurosurgery, and Psychiatry*, 47(4), 377–385.
- Laramée, M. E., Rockland, K. S., Prince, S., Bronchti, G., & Boire, D. (2013). Principal component and cluster analysis of layer V pyramidal cells in visual and non-visual cortical areas projecting to the primary visual cortex of the mouse. *Cerebral Cortex*, 23(3), 714–728.
- Larkum, M. E., Senn, W., & Luscher, H. R. (2004). Top-down dendritic input increases the gain of layer 5 pyramidal neurons. *Cerebral Cortex*, 14(10), 1059–1070. doi:10.1093/cercor/bhh065.
- Larkum, M. E., Zhu, J. J., & Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398(6725), 338–341.
- Lashley, K. S. (1943). Studies of cerebral function in learning XII. Loss of the maze habit after occipital lesions in blind rats. *Journal of Comparative Neurology*, 79, 431–462.
- Lau, H. C. (2008). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research*, 168, 35–48. doi:10.1016/S0079-6123(07)68004-2.
- Laurent, G., MacLeod, K., Stopfer, M., & Wehr, M. (1998). Spatiotemporal structure of olfactory inputs to the mushroom bodies. *Learning & Memory*, 5(1–2), 124–132.
- Laureys, S., Pellas, F., Van Eeckhout, P., Ghorbel, S., Schnakers, C., Perrin, F., et al. (2005). The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless? *Progress in Brain Research*, 150, 495–511. doi:10.1016/S0079-6123(05)50034-7.
- Le Bon-Jego, M., & Yuste, R. (2007). Persistently active, pacemaker-like neurons in neocortex. *Frontiers in Neuroscience*, 1(1), 123–129. doi:10.3389/neuro.01.1.1.009.2007.
- LeDoux, J. E. (1994). Emotion, memory and the brain. *Scientific American*, 270(6), 50–57.
- Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, 36(6), 1183–1194.
- Lee, C. C., & Winer, J. A. (2011). Convergence of thalamic and cortical pathways in cat auditory cortex. *Hearing Research*, 274(1–2), 85–94. doi:10.1016/j.heares.2010.05.008.

- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 20(7), 1434–1448.
- Leech, G. N. (1977). *Semantics*. Harmondsworth: Penguin Books.
- Lefort, S., Tomm, C., Floyd Sarria, J. C., & Petersen, C. C. (2009). The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron*, 61(2), 301–316. doi:10.1016/j.neuron.2008.12.020.
- Lehrer, J. (2007). *Proust was a neuroscientist*. Boston: Houghton Mifflin.
- Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science*, 317(5841), 1096–1099. doi:10.1126/science.1143439.
- Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, 379(6565), 549–553. doi:10.1038/379549a0.
- Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences*, 3(7), 254–264.
- Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., & Moser, M. B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 309(5734), 619–623.
- Lev-Ram, V., Mehta, S. B., Kleinfeld, D., & Tsien, R. Y. (2003). Reversing cerebellar long-term depression. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26), 15989–15993. doi:10.1073/pnas.2636935100.
- Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64, 354–361.
- Levine, J. (2003). Experience and representation. In Q. Smith & A. Jokic (Eds.), *Consciousness: new philosophical perspectives* (pp. 57–76). Oxford: Oxford University Press.
- Lewis, D. (1980a). Psychophysical and theoretical identifications. In N. Block (Ed.), *Readings in philosophy of psychology* (pp. 207–221). Cambridge, MA: Harvard University Press.
- Lewis, D. (1980b). Mad pain and martial pain. In N. Block (Ed.), *Readings in philosophy of psychology* (pp. 216–222). Cambridge, MA: Harvard University Press.
- Lewis, M. (2003). The emergence of consciousness and its role in human development. *Annals of the New York Academy of Sciences*, 1001, 104–133.
- Li, J., Bickford, M. E., & Guido, W. (2003). Distinct firing properties of higher order thalamic relay neurons. *Journal of Neurophysiology*, 90(1), 291–299.
- Li, L. Y., Li, Y. T., Zhou, M., Tao, H. W., & Zhang, L. I. (2013). Intracortical multiplication of thalamocortical signals in mouse auditory cortex. *Nature Neuroscience*, 16(9), 1179–1181. doi:10.1038/nn.3493.
- Li, W., Lopez, L., Osher, J., Howard, J. D., Parrish, T. B., & Gottfried, J. A. (2010). Right orbitofrontal cortex mediates conscious olfactory perception. *Psychological Science*, 21(10), 1454–1463. doi:10.1177/0956797610382121.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain*, 106(Pt 3), 623–642.
- Lima, B., Singer, W., Chen, N. H., & Neuenschwander, S. (2010). Synchronization dynamics in response to plaid stimuli in monkey V1. *Cerebral Cortex*, 20(7), 1556–1573. doi:10.1093/cercor/bhp218.
- Linden, J. F., & Schreiner, C. E. (2003). Columnar transformations in auditory cortex? A comparison to visual and somatosensory cortices. *Cerebral Cortex*, 13(1), 83–89.
- Linkenkaer-Hansen, K., Nikulin, V. V., Palva, S., Ilmoniemi, R. J., & Palva, J. M. (2004). Prestimulus oscillations enhance psychophysical performance in humans. *Journal of Neuroscience*, 24(45), 10186–10190. doi:10.1523/JNEUROSCI.2584-04.2004.
- Lisman, J. E., & Idiart, M. A. (1995). Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science*, 267(5203), 1512–1515.
- Lissauer, H. (1890). Ein Fall von Seelendblindheit nebst einem Beitrag zur Theorie derselben (A case of visual agnosia with a contribution to theory). *Archiv für Psychiatrie und Nervenkrankheiten*, 21, 222–270.
- Little, W. A. (1974). The existence of persistent states in the brain. *Mathematical Biosciences*, 19, 101–120.

- Llinás, R. (2001). *I of the vortex*. Cambridge, MA: MIT Press.
- Llinás, R., Lang, E. J., & Welsh, J. P. (1997). The cerebellum, LTD, and memory: alternative views. *Learning & Memory*, 3(6), 445–455.
- Llinás, R., & Ribary, U. (1993). Coherent 40-Hz oscillation characterizes dream state in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 90(5), 2078–2081.
- Llinás, R., & Ribary, U. (2001). Consciousness and the brain: the thalamocortical dialogue in health and disease. *Annals of the New York Academy of Sciences*, 929, 166–175.
- Llinás, R. R. (2011). Cerebellar motor learning versus cerebellar motor timing: the climbing fibre story. *Journal of Physiology*, 589(Pt 14), 3423–3432. doi:10.1113/jphysiol.2011.207464.
- Lloyd, D. M., Shore, D. I., Spence, C., & Calvert, G. A. (2003). Multisensory representation of limb position in human premotor cortex. *Nature Neuroscience*, 6(1), 17–18.
- Locke, J. (1667). *An essay concerning human understanding* (5th ed.). London: Collins Press.
- Lodge, D. (2001). *Thinks....* London: Penguin books.
- Lopes da Silva, F. H. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalography and Clinical Neurophysiology*, 79, 81–93.
- Lopes da Silva, F. H. (2004). Contribution to a neurophysiology of consciousness. *Supplements to Clinical Neurophysiology*, 57(Suppl), 645–655.
- Losavio, B. E., Liang, Y., Santamaría-Pang, A., Kakadiaris, I. A., Colbert, C. M., & Saggau, P. (2008). Live neuron morphology automatically reconstructed from multiphoton and confocal imaging data. *Journal of Neurophysiology*, 100(4), 2422–2429.
- Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, 29(1), 145–156.
- Luczak, A., Bartho, P., & Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, 62(3), 413–425. doi:10.1016/j.neuron.2009.03.014.
- Lulé, D., Zickler, C., Hacker, S., Bruno, M. A., Demertzi, A., Pellas, F., et al. (2009). Life can be worth living in locked-in syndrome. *Progress in Brain Research*, 177, 339–351. doi:10.1016/S0079-6123(09)17723-3.
- Lund, J. S. (1988). Anatomical organization of macaque monkey striate visual cortex. *Annual Review of Neuroscience*, 11, 253–288. doi:10.1146/annurev.ne.11.030188.001345.
- Lund, J. S., Angelucci, A., & Bressloff, P. C. (2003). Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cerebral Cortex*, 13(1), 15–24.
- Luscher, C., Nicoll, R. A., Malenka, R. C., & Muller, D. (2000). Synaptic plasticity and dynamic modulation of the postsynaptic membrane. *Nature Neuroscience*, 3(6), 545–550. doi:10.1038/75714.
- Lycan, W. G. (1987). *Consciousness*. Cambridge, MA: MIT Press.
- Lycan, W. G. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.
- Lycan, W. G. (2001). The case for phenomenal externalism. *Noûs*, 35(Suppl), 17–35.
- Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in Neurosciences*, 28(5), 264–271.
- Macaluso, E., Frith, C. D., & Driver, J. (2002). Crossmodal spatial influences of touch on extrastriate visual areas take current gaze direction into account. *Neuron*, 34(4), 647–658.
- Macaluso, E., & Maravita, A. (2010). The representation of space near the body through touch and vision. *Neuropsychologia*, 48(3), 782–795. doi:10.1016/j.neuropsychologia.2009.10.010.
- MacKay, D. M. (1962). Theoretical models of space perception. In C. A. Muses (Ed.), *Aspects of the theory of artificial intelligence* (pp. 83–104). New York: Plenum Press.
- MacKay, D. M. (1965). A mind's eye view of the brain. *Progress in Brain Research*, 17, 321–332.
- MacKay, D. M. (1973). Visual stability and voluntary eye movements. In R. Jung (Ed.), *Handbook of sensory physiology* (Vol. VII/3A, pp. 307–331). Berlin: Springer.
- Malenka, R. C., & Nicoll, R. A. (1999). Long-term potentiation—a decade of progress? *Science*, 285(5435), 1870–1874. doi:7846 [pii].

- Malmierca, M. S., Izquierdo, M. A., Cristaudo, S., Hernandez, O., Perez-Gonzalez, D., Covey, E., et al. (2008). A discontinuous tonotopic organization in the inferior colliculus of the rat. *Journal of Neuroscience*, 28(18), 4767–4776.
- Malnic, B., Hirono, J., Sato, T., & Buck, L. B. (1999). Combinatorial receptor codes for odors. *Cell*, 96(5), 713–723.
- Maquet, P., Peters, J., Aerts, J., Delfiore, G., Degueldre, C., Luxen, A., et al. (1996). Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. *Nature*, 383(6596), 163–166. doi:10.1038/383163a0.
- Marblestone, A. H., Zamft, B. M., Maguire, Y. G., Shapiro, M. G., Cybulski, T. R., Glaser, J. I., et al. (2013). Physical principles for scalable neural recording. *Frontiers in Computational Neuroscience*, 7, 137. doi:10.3389/fncom.2013.00137.
- Marcel, A. J. (1983). Conscious and unconscious perception: an approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, 15(2), 238–300.
- Marcus, G. F. (2001). *The algebraic mind: integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.
- Maren, S., & Quirk, G. J. (2004). Neuronal signalling of fear memory. *Nature Reviews Neuroscience*, 5(11), 844–852.
- Marin, O., Smeets, W. J., & Gonzalez, A. (1998). Evolution of the basal ganglia in tetrapods: a new perspective based on recent studies in amphibians. *Trends in Neurosciences*, 21(11), 487–494.
- Markowitsch, H. J., & Kessler, J. (2000). Massive impairment in executive functions with partial preservation of other cognitive functions: the case of a young patient with severe degeneration of the prefrontal cortex. *Experimental Brain Research*, 133(1), 94–102.
- Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology*, 202(2), 437–470.
- Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 262(841), 23–81.
- Marr, D. (1982). *Vision*. New York: W.H. Freeman.
- Martinez, D. (2005). Detailed and abstract phase-locked attractor network models of early olfactory systems. *Biological Cybernetics*, 93(5), 355–365. doi:10.1007/s00422-005-0010-3.
- Mashour, G. A., Walker, E. E., & Martuza, R. L. (2005). Psychosurgery: past, present, and future. *Brain Research. Brain Research Reviews*, 48(3), 409–419.
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., & Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science*, 309(5744), 2228–2232. doi:10.1126/science.1117256.
- Massimini, M., Huber, R., Ferrarelli, F., Hill, S., & Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *Journal of Neuroscience*, 24(31), 6862–6870. doi:10.1523/JNEUROSCI.1318-04.2004.
- Mateen, F. J., & Josephs, K. A. (2008). Sudden behavioral disturbance in a man with a lesion in the nucleus accumbens. *Journal of Neurology*, 255(11), 1834–1835. doi:10.1007/s00415-008-0995-3.
- Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *Journal of Neuroscience*, 29(9), 2725–2732. doi:10.1523/JNEUROSCI.3963-08.2009.
- Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10, 363–401.
- Mayr, E. (1988). *Toward a new philosophy of biology: observations of an evolutionist*. Cambridge, MA: Harvard University Press.
- Mazzoni, P., Andersen, R. A., & Jordan, M. I. (1991). A more biologically plausible learning rule for neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, 88(10), 4433–4437.
- McClelland, J., Rumelhart, D., & the PDP-Research-Group. (1986). *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.

- McCormick, D. A., & Wang, Z. (1991). Serotonin and noradrenaline excite GABAergic neurones of the guinea-pig and cat nucleus reticularis thalami. *Journal of Physiology*, 442, 235–255.
- McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- McGaugh, J. L. (2000). Memory—a century of consolidation. *Science*, 287(5451), 248–251.
- McGinn, C. (1989). Can we solve the mind-body problem? *Mind*, 98, 349–366.
- McGuire, B. A., Gilbert, C. D., Rivlin, P. K., & Wiesel, T. N. (1991). Targets of horizontal connections in macaque primary visual cortex. *Journal of Comparative Neurology*, 305(3), 370–392. doi:10.1002/cne.903050303.
- McKeefry, D. J., & Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*, 120, 2229–2242.
- McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the “cognitive map.” *Nature Reviews Neuroscience*, 7(8), 663–678.
- Medalla, M., & Barbas, H. (2010). Anterior cingulate synapses in prefrontal areas 10 and 46 suggest differential influence in cognitive control. *Journal of Neuroscience*, 30(48), 16068–16081. doi:10.1523/JNEUROSCI.1773-10.2010.
- Meeren, H. K., Pijn, J. P., Van Luijtelaar, E. L., Coenen, A. M., & Lopes da Silva, F. H. (2002). Cortical focus drives widespread corticothalamic networks during spontaneous absence seizures in rats. *Journal of Neuroscience*, 22(4), 1480–1495.
- Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30(1), 63–81, discussion 81–134. doi:10.1017/S0140525X07000891.
- Merleau-Ponty, M. (1945). *Phenomenology of perception* (translated by C. Smith, 1958). London: Routledge & Kegan Paul.
- Merzenich, M. M., & Brugge, J. F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Research*, 50(2), 275–296.
- Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology*, 10(4), 309–325. doi:10.1002/ana.410100402.
- Mesulam, M. M. (1998). From sensation to cognition. *Brain*, 121(Pt 6), 1013–1052.
- Metin, C., & Frost, D. O. (1989). Visual responses of neurons in somatosensory cortex of hamsters with experimentally induced retinal projections to somatosensory thalamus. *Proceedings of the National Academy of Sciences of the United States of America*, 86(1), 357–361.
- Metzinger, T. (2000). The subjectivity of subjective experience: a representationalist analysis of the first-person perspective. In T. Metzinger (Ed.), *Neural correlates of consciousness—empirical and conceptual questions* (pp. 285–306). Cambridge, MA: MIT Press.
- Metzinger, T. (2008). Empirical perspectives from the self-model theory of subjectivity: a brief summary with examples. *Progress in Brain Research*, 168, 215–245. doi:10.1016/S0079-6123(07)68018-2.
- Meyer, K. (2011). Primary sensory cortices, top-down projections and conscious experience. *Progress in Neurobiology*, 94(4), 408–417. doi:10.1016/j.pneurobio.2011.05.010.
- Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, 100(3), 1407–1419. doi:10.1152/jn.90248.2008.
- Miconi, T., & VanRullen, R. (2010). The gamma slideshow: object-based perceptual cycles in a model of the visual cortex. *Frontiers in Human Neuroscience*, 4, 205. doi:10.3389/fnhum.2010.00205.
- Milad, M. R., & Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*, 420(6911), 70–74. doi:10.1038/nature01138.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, 16(16), 5154–5167.
- Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.

- Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review*, 81(6), 521–535.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA.: MIT Press.
- Mithen, S. (1999). Handaxes and Ice Age carvings: hard evidence for the evolution of consciousness. In S. R. Hameroff, A. W. Kaszniak, & D. J. Chalmers (Eds.), *Toward a science of consciousness III: the third Tucson discussion and debates* (Vol. 3, pp. 281–296). Cambridge, MA: MIT Press.
- Mittelstaedt, M. L., & Mittelstaedt, H. (2001). Idiothetic navigation in humans: estimation of path length. *Experimental Brain Research*, 139(3), 318–332.
- Mombaerts, P. (2004). Odorant receptor gene choice in olfactory sensory neurons: the one receptor—one neuron hypothesis revisited. *Current Opinion in Neurobiology*, 14(1), 31–36. doi:10.1016/j.conb.2004.01.014.
- Montaser-Kouhsari, L., & Rajimehr, R. (2004). Attentional modulation of adaptation to illusory lines. *Journal of Vision*, 4(6), 434–444. doi:10.1167/4.6.3.
- Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., et al. (2010). Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine*, 362(7), 579–589. doi:10.1056/Nejmoa0905370.
- Montijn, J. S., Vinck, M., & Pennartz, C. M. (2014). Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation. *Frontiers in Computational Neuroscience*, 8, 58. doi:10.3389/fncom.2014.00058.
- Moroni, F., Nobili, L., Curcio, G., De Carli, F., Fratello, F., Marzano, C., et al. (2007). Sleep in the human hippocampus: a stereo-EEG study. *PLoS ONE*, 2(9), e867. doi:10.1371/journal.pone.0000867.
- Morris, R. G., Anderson, E., Lynch, G. S., & Baudry, M. (1986). Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP5. *Nature*, 319(6056), 774–776. doi:10.1038/319774a0.
- Morrison, S. F., Nakamura, K., & Madden, C. J. (2008). Central control of thermogenesis in mammals. *Experimental Physiology*, 93(7), 773–797.
- Morita, M., & Suemitsu, A. (2002). Computational modeling of pair-association memory in inferior temporal cortex. *Brain Research. Cognitive Brain Research*, 13(2), 169–178.
- Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., & Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current Opinion in Neurobiology*, 16(2), 179–190. doi:10.1016/j.conb.2006.03.013.
- Muetzelfeldt, L., Kamboj, S. K., Rees, H., Taylor, J., Morgan, C. J., & Curran, H. V. (2008). Journey through the K-hole: phenomenological aspects of ketamine use. *Drug and Alcohol Dependence*, 95(3), 219–229. doi:10.1016/j.drugalcdep.2008.01.024.
- Müller, J. (1838). Handbuch der Physiologie des Menschen für Vorlesungen (translated as *Elements of physiology* by W. M. Baly, 1843, Philadelphia: Lea and Blanchard). Coblenz: J. Hölscher.
- Müller, M. K., Tremer, M., Bodenstein, C., & Würtz, R. P. (2013). Learning invariant face recognition from examples. *Neural Networks*, 41, 137–146.
- Müller, N. G., & Knight, R. T. (2006). The functional neuroanatomy of working memory: contributions of human brain lesion studies. *Neuroscience*, 139(1), 51–58. doi:10.1016/j.neuroscience.2005.09.018.
- Murakami, I., Kitaoka, A., & Ashida, H. (2006). A positive correlation between fixation instability and the strength of illusory motion in a static display. *Vision Research*, 46(15), 2421–2431. doi:10.1016/j.visres.2006.01.030.
- Murphy, C., Cain, W. S., & Bartoshuk, L. M. (1977). Mutual action of taste and olfaction. *Sensory Processes*, 1(3), 204–211.
- Murray, E. A., Bussey, T. J., & Saksida, L. M. (2007). Visual perception and memory: a new view of medial temporal lobe function in primates and rodents. *Annual Review of Neuroscience*, 30, 99–122. doi:10.1146/annurev.neuro.29.051605.113046.
- Murthy, V. N., & Fetz, E. E. (1996a). Oscillatory activity in sensorimotor cortex of awake monkeys: synchronization of local field potentials and relation to behavior. *Journal of Neurophysiology*, 76(6), 3949–3967.

- Murthy, V. N., & Fetz, E. E. (1996b). Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys. *Journal of Neurophysiology*, 76(6), 3968–3982.
- Myles, P. S., Leslie, K., McNeil, J., Forbes, A., & Chan, M. T. (2004). Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial. *Lancet*, 363(9423), 1757–1763. doi:10.1016/S0140-6736(04)16300-9.
- Nadasdy, Z. (2010). Binding by asynchrony: the neuronal phase code. *Frontiers in Neuroscience*, 4. doi:10.3389/fnins.2010.00051.
- Nadasdy, Z., Hirase, H., Czurko, A., Csicsvari, J., & Buzsaki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, 19(21), 9497–9507.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450.
- Nagel, T. (1980). Armstrong on the mind. In N. Block (Ed.), *Readings in philosophy of psychology* (pp. 200–206). Cambridge, MA: Harvard University Press.
- Nagel, T. (1986). *The view from nowhere*. New York: Oxford University Press.
- Nakamura, R. K., & Mishkin, M. (1986). Chronic “blindness” following lesions of nonvisual cortex in the monkey. *Experimental Brain Research*, 63(1), 173–184.
- Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13(12), 1362–1368.
- Nauta, H. J. (1979). A proposed conceptual reorganization of the basal ganglia and telencephalon. *Neuroscience*, 4(12), 1875–1881.
- Neggers, S. F., Van der Lubbe, R. H., Ramsey, N. F., & Postma, A. (2006). Interactions between ego- and allocentric neuronal representations of space. *NeuroImage*, 31(1), 320–331.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century Crofts.
- Nelson, R. J. (1996). Interactions between motor commands and somatic perception in sensorimotor cortex. *Current Opinion in Neurobiology*, 6(6), 801–810.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newsome, W. T., & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, 8(6), 2201–2211.
- Nichols, M. J., & Newsome, W. T. (2002). Middle temporal visual area microstimulation influences veridical judgments of motion direction. *Journal of Neuroscience*, 22(21), 9530–9540.
- Nicola, S. M., Surmeier, J., & Malenka, R. C. (2000). Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annual Review of Neuroscience*, 23, 185–215. doi:10.1146/annurev.neuro.23.1.185.
- Nieuwenhuys, R., Voogd, J., & Van Huijzen, C. (1988). *The human central nervous system: a synopsis and atlas* (3rd ed.). Berlin: Springer-Verlag.
- Noppeney, U., Patterson, K., Tyler, L. K., Moss, H., Stamatakis, E. A., Bright, P., et al. (2007). Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. *Brain*, 130(Pt 4), 1138–1147. doi:10.1093/brain/awl344.
- Noreika, V., Jylhankangas, L., Moro, L., Valli, K., Kaskinoro, K., Aantaa, R., et al. (2011). Consciousness lost and found: subjective experiences in an unresponsive state. *Brain and Cognition*, 77(3), 327–334. doi:10.1016/j.bandc.2011.09.002.
- Norris, S. A., Hathaway, E. N., Taylor, J. A., & Thach, W. T. (2011). Cerebellar inactivation impairs memory of learned prism gaze-reach calibrations. *Journal of Neurophysiology*, 105(5), 2248–2259. doi:10.1152/jn.01009.2010.
- Nothdurft, H. C. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception & Psychophysics*, 52, 355–375.
- Nowlan, S. J. (1990). *Competing Experts: an experimental investigation of associative mixture models*. Technical Report CRG-TR-90-5, University of Toronto.
- O'Brien, G., & Opie, J. (1999). A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences*, 22, 127–148.

- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1), 171–175. doi:0006-8993(71)90358-1 [pii].
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3), 317–330.
- O'Neil, E. B., Protzner, A. B., McCormick, C., McLean, D. A., Poppenk, J., Cate, A. D., et al. (2012). Distinct patterns of functional and effective connectivity between perirhinal cortex and other cortical regions in recognition memory and perceptual discrimination. *Cerebral Cortex*, 22(1), 74–85. doi:10.1093/cercor/bhr075.
- O'Regan, J. K., Myin, E., & Noë, A. (2005). Sensory consciousness explained (better) in terms of “corporality” and “alerting capacity.” *Phenomenology and the Cognitive Sciences*, 4, 369–387.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 939–973.
- Obrietan, K., & van den Pol, A. N. (1995). GABA neurotransmission in the hypothalamus: developmental reversal from Ca²⁺ elevating to depressing. *Journal of Neuroscience*, 15(7 Pt 1), 5065–5077.
- Olausson, H., Lamarre, Y., Backlund, H., Morin, C., Wallin, B. G., Starck, G., et al. (2002). Unmyelinated tactile afferents signal touch and project to insular cortex. *Nature Neuroscience*, 5(9), 900–904. doi:10.1038/nn896.
- Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex*, 10(3), 206–219.
- Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, G. Maxwell, & M. Scriven (Eds.), *Concepts, theories, and the mind–body problem* (pp. 3–36). Minneapolis: University of Minnesota Press.
- Oram, M. W., Wiener, M. C., Lestienne, R., & Richmond, B. J. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *Journal of Neurophysiology*, 81(6), 3021–3033.
- Orsini, C. A., Kim, J. H., Knapska, E., & Maren, S. (2011). Hippocampal and prefrontal projections to the basal amygdala mediate contextual regulation of fear after extinction. *Journal of Neuroscience*, 31(47), 17269–17277. doi:10.1523/JNEUROSCI.4095-11.2011.
- Oswald, A. M., & Reyes, A. D. (2008). Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex. *Journal of Neurophysiology*, 99(6), 2998–3008. doi:10.1152/jn.01160.2007.
- Oudiette, D., Dealberto, M. J., Uguzzioni, G., Golmard, J. L., Merino-Andreu, M., Tafti, M., et al. (2012). Dreaming without REM sleep. *Consciousness and Cognition*, 21(3), 1129–1140. doi:10.1016/j.concog.2012.04.010.
- Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science*, 313(5792), 1402. doi:10.1126/science.1130197.
- Pakkenberg, B., & Gundersen, H. J. (1997). Neocortical neuron number in humans: effect of sex and age. *Journal of Comparative Neurology*, 384(2), 312–320.
- Pallis, C. A. (1955). Impaired identification of faces and places with agnosia for colours; report of a case due to cerebral embolism. *Journal of Neurology, Neurosurgery, and Psychiatry*, 18(3), 218–224.
- Palmer, S. E. (1999). Color, consciousness, and the isomorphism constraint. *Behavioral and Brain Sciences*, 22(6), 923–943, discussion 944–989.
- Palva, S., Linkenkaer-Hansen, K., Naatanen, R., & Palva, J. M. (2005). Early neural correlates of conscious somatosensory perception. *Journal of Neuroscience*, 25(21), 5248–5258. doi:10.1523/JNEUROSCI.0141-05.2005.
- Palva, S., & Palva, J. M. (2007). New vistas for alpha-frequency band oscillations. *Trends in Neurosciences*, 30(4), 150–158. doi:10.1016/j.tins.2007.02.001.
- Panayiotopoulos, C. P., Obeid, T., & Waheed, G. (1989). Differentiation of typical absence seizures in epileptic syndromes: a video EEG study of 224 seizures in 20 patients. *Brain*, 112(Pt 4), 1039–1056.
- Papoulis, A. (1991). *Probability, random variables and stochastic processes*. New York: McGraw-Hill.
- Parkinson, J. A., Crofts, H. S., McGuigan, M., Tomic, D. L., Everitt, B. J., & Roberts, A. C. (2001). The role of the primate amygdala in conditioned reinforcement. *Journal of Neuroscience*, 21(19), 7770–7780.

- Parsons, L. M., Petacchi, A., Schmahmann, J. D., & Bower, J. M. (2009). Pitch discrimination in cerebellar patients: evidence for a sensory deficit. *Brain Research*, 1303, 84–96. doi:10.1016/j.brainres.2009.09.052.
- Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292(5516), 510–512. doi:10.1126/science.1057099.
- Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321(5894), 1322–1327. doi:10.1126/science.1159775.
- Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, 439(7078), 865–870. doi:10.1038/nature04490.
- Pauli, W. M., Hazy, T. E., & O'Reilly, R. C. (2012). Expectancy, ambiguity, and behavioral flexibility: separable and complementary roles of the orbital frontal cortex and amygdala in processing reward expectancies. *Journal of Cognitive Neuroscience*, 24(2), 351–366. doi:10.1162/jocn_a_00155.
- Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *Journal of Neuroscience*, 9(8), 2907–2918.
- Peigneux, P., Orban, P., Balteau, E., Degueldre, C., Luxen, A., Laureys, S., et al. (2006). Offline persistence of memory-related cerebral activity during active wakefulness. *PLoS Biology*, 4(4), e100. doi:10.1371/journal.pbio.0040100.
- Penfield, W. (1958). Some mechanisms of consciousness discovered during electrical stimulation of the brain. *Proceedings of the National Academy of Sciences of the United States of America*, 44(2), 51–66.
- Penfield, W. (1975). *The mystery of the mind*. Princeton, NJ: Princeton University Press.
- Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man*. New York: MacMillan.
- Pennartz, C. M., Berke, J. D., Graybiel, A. M., Ito, R., Lansink, C. S., van der Meer, M., et al. (2009). Corticostriatal interactions during learning, memory processing, and decision making. *Journal of Neuroscience*, 29(41), 12831–12838. doi:10.1523/JNEUROSCI.3177-09.2009.
- Pennartz, C. M., Boeijinga, P. H., Kitai, S. T., & Lopes da Silva, F. H. (1991). Contribution of NMDA receptors to postsynaptic potentials and paired-pulse facilitation in identified neurons of the rat nucleus accumbens in vitro. *Experimental Brain Research*, 86(1), 190–198.
- Pennartz, C. M., de Jeu, M. T., Bos, N. P., Schaap, J., & Geurtsen, A. M. (2002). Diurnal modulation of pacemaker potentials and calcium current in the mammalian circadian clock. *Nature*, 416(6878), 286–290.
- Pennartz, C. M., Dolleman-Van der Weel, M. J., Kitai, S. T., & Lopes da Silva, F. H. (1992). Presynaptic dopamine D1 receptors attenuate excitatory and inhibitory limbic inputs to the shell region of the rat nucleus accumbens studied in vitro. *Journal of Neurophysiology*, 67(5), 1325–1334.
- Pennartz, C. M., Groenewegen, H. J., & Lopes da Silva, F. H. (1994). The nucleus accumbens as a complex of functionally distinct neuronal ensembles: an integration of behavioural, electrophysiological and anatomical data. *Progress in Neurobiology*, 42(6), 719–761.
- Pennartz, C. M., Ito, R., Verschure, P. F., Battaglia, F. P., & Robbins, T. W. (2011). The hippocampal–striatal axis in learning, prediction and goal-directed behavior. *Trends in Neurosciences*, 34(10), 548–559. doi:10.1016/j.tins.2011.08.001.
- Pennartz, C. M., Lee, E., Verheul, J., Lipa, P., Barnes, C. A., & McNaughton, B. L. (2004). The ventral striatum in off-line processing: ensemble reactivation during sleep and modulation by hippocampal ripples. *Journal of Neuroscience*, 24(29), 6446–6456. doi:10.1523/JNEUROSCI.0575-04.2004.
- Pennartz, C. M., Uylings, H. B., Barnes, C. A., & McNaughton, B. L. (2002). Memory reactivation and consolidation during sleep: from cellular mechanisms to human performance. *Progress in Brain Research*, 138, 143–166. doi:10.1016/S0079-6123(02)38076-2.
- Pennartz, C. M., van Wingerden, M., & Vinck, M. (2011). Population coding and neural rhythmicity in the orbitofrontal cortex. *Annals of the New York Academy of Sciences*, 1239(1), 149–161. doi:10.1111/j.1749-6632.2011.06296.x.
- Pennartz, C. M. A. (1996). The ascending neuromodulatory systems in learning by reinforcement: comparing computational conjectures with experimental findings. *Brain Research. Brain Research Reviews*, 21(3), 219–245.

- Pennartz, C. M. A. (1997). Reinforcement learning by Hebbian synapses with adaptive thresholds. *Neuroscience*, 81(2), 303–319.
- Pennartz, C. M. A. (2009). Identification and integration of sensory modalities: neural basis and relation to consciousness. *Consciousness and Cognition*, 18(3), 718–739. doi:10.1016/j.concog.2009.03.003.
- Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., & Laurent, G. (2002). Oscillations and sparsening of odor representations in the mushroom body. *Science*, 297(5580), 359–365. doi:10.1126/science.1070502.
- Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 335(1273), 23–30. doi:10.1098/rstb.1992.0003.
- Person, A. L., & Raman, I. M. (2012). Synchrony and neural coding in cerebellar circuits. *Frontiers in Neural Circuits*, 6, 97. doi:10.3389/fncir.2012.00097.
- Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., & Frith, C. D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. *Neuron*, 59(4), 561–567. doi:10.1016/j.neuron.2008.07.005.
- Peterson, M. A., & Enns, J. T. (2005). The edge complex: implicit memory for figure assignment in shape perception. *Perception & Psychophysics*, 67, 727–740.
- Petreanu, L., Huber, D., Sobczyk, A., & Svoboda, K. (2007). Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neuroscience*, 10(5), 663–668. doi:10.1038/nn1891.
- Petreanu, L., Mao, T., Sternson, S. M., & Svoboda, K. (2009). The subcellular organization of neocortical excitatory connections. *Nature*, 457(7233), 1142–1145. doi:10.1038/nature07709.
- Petrides, M., & Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *Journal of Comparative Neurology*, 228(1), 105–116. doi:10.1002/cne.902280110.
- Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., & Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, 12(7), 919–926. doi:10.1038/nn.2337.
- Pezzulo, G., Van der Meer, M. A. A., Lansink, C. S., & Pennartz, C. M. A. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends in Cognitive Sciences*, 18, 647–657.
- Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447), 74–79. doi:10.1038/nature12112.
- Pfurtscheller, G. (2001). Functional brain imaging based on ERD/ERS. *Vision Research*, 41(10–11), 1257–1260.
- Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, 48(2), 175–187.
- Pitkänen, A., Savander, V., & LeDoux, J. E. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends in Neurosciences*, 20(11), 517–523. doi:S0166223697011259 [pii].
- Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology*, 47, 42–51.
- Plant, T. M. (2012). A comparison of the neuroendocrine mechanisms underlying the initiation of the preovulatory LH surge in the human, Old World monkey and rodent. *Frontiers in Neuroendocrinology*, 33(2), 160–168. doi:10.1016/j.yfrne.2012.02.002.
- Plenz, D. (2003). When inhibition goes incognito: feedback interaction between spiny projection neurons in striatal function. *Trends in Neurosciences*, 26(8), 436–443. doi:10.1016/S0166-2236(03)00196-6.
- Plotnik, J. M., de Waal, F. B., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences of the United States of America*, 103(45), 17053–17057. doi:10.1073/pnas.0608062103.
- Plum, F., Schiff, N., Ribary, U., & Llinás, R. (1998). Coordinated expression in chronically unconscious persons. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 353(1377), 1929–1933. doi:10.1098/rstb.1998.0345.

- Pockett, S., & Holmes, M. D. (2009). Intracranial EEG power spectra and phase synchrony during consciousness and unconsciousness. *Consciousness and Cognition*, 18(4), 1049–1055. doi:10.1016/j.concog.2009.08.010.
- Poe, G. R., Nitz, D. A., McNaughton, B. L., & Barnes, C. A. (2000). Experience-dependent phase-reversal of hippocampal neuron firing during REM sleep. *Brain Research*, 855(1), 176–180.
- Polanyi, D. (1966). *The tacit dimension*. Garden City, NY: Doubleday.
- Pollatos, O., Kirsch, W., & Schandry, R. (2005). Brain structures involved in interoceptive awareness and cardioafferent signal processing: a dipole source localization study. *Human Brain Mapping*, 26(1), 54–64. doi:10.1002/hbm.20121.
- Pollen, D. A. (1995). Cortical areas in visual awareness. *Nature*, 377(6547), 293–295. doi:10.1038/377293b0.
- Pollen, D. A. (1999). On the neural correlates of visual perception. *Cerebral Cortex*, 9(1), 4–19.
- Pollen, D. A. (2011). On the emergence of primary visual perception. *Cerebral Cortex*, 21(9), 1941–1953. doi:10.1093/cercor/bhq285.
- Polonsky, A., Blake, R., Braun, J., & Heeger, D. J. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature Neuroscience*, 3(11), 1153–1159. doi:10.1038/80676.
- Pomerleau, D. A., Gowdy, J., & Thorpe, C. E. (1991). Combining artificial neural networks and symbolic processing for autonomous robot guidance. *Engineering Applications of Artificial Intelligence*, 4, 279–285.
- Popper, K., & Eccles, J. C. (1977). *The self and its brain*. Berlin: Springer Verlag.
- Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 91(16), 7398–7403.
- Postle, B. R., Berger, J. S., & D'Esposito, M. (1999). Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proceedings of the National Academy of Sciences of the United States of America*, 96(22), 12959–12964.
- Pouget, A., Deneve, S., & Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3(9), 741–747.
- Powell, T. P., & Mountcastle, V. B. (1959). Some aspects of the functional organization of the cortex of the postcentral gyrus of the monkey: a correlation of findings obtained in a single unit analysis with cytoarchitecture. *Bulletin of the Johns Hopkins Hospital*, 105, 133–162.
- Prager-Khoutorsky, M., & Bourque, C. W. (2010). Osmosensation in vasopressin neurons: changing actin density to optimize function. *Trends in Neurosciences*, 33(2), 76–83. doi:10.1016/j.tins.2009.11.004.
- Ptito, A., Crane, J., Leonard, G., Amsel, R., & Caramanos, Z. (1995). Visual–spatial localization by patients with frontal-lobe lesions invading or sparing area 46. *Neuroreport*, 6(13), 1781–1784.
- Putnam, F. W. (1993). Dissociative disorders in children: behavioral profiles and problems. *Child Abuse & Neglect*, 17(1), 39–45.
- Putnam, H. (1975a). The meaning of “meaning.” In K. Gunderson (Ed.), *Language, mind, and knowledge: Minnesota studies in the philosophy of science* (Vol. VII, pp. 131–193). Minneapolis: University of Minnesota Press.
- Putnam, H. (1975b). *Mind, language and reality*. Cambridge: Cambridge University Press.
- Pylyshyn, Z. (1980). Computation and cognition: issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3, 111–169.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3), 341–365, discussion 366–423.
- Querfurth, H. W., & LaFerla, F. M. (2010). Alzheimer's disease. *New England Journal of Medicine*, 362(4), 329–344. doi:10.1056/NEJMra0909142.
- Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nature Reviews Neuroscience*, 13(8), 587–597. doi:10.1038/nrn3251.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102–1107. doi:10.1038/nature03687.

- Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage*, 37(4), 1083–1090; discussion 1097–1089. doi:10.1016/j.neuroimage.2007.02.041.
- Raij, T. T., Numminen, J., Narvanen, S., Hiltunen, J., & Hari, R. (2005). Brain correlates of subjective reality of physically and psychologically induced pain. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6), 2147–2151. doi:10.1073/pnas.0409542102.
- Rainer, G., Rao, S. C., & Miller, E. K. (1999). Prospective coding for objects in primate prefrontal cortex. *Journal of Neuroscience*, 19(13), 5493–5505.
- Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science*, 277(5328), 968–971.
- Ramachandran, V. S. (2005). Plasticity and functional recovery in neurology. *Clinical Medicine (London, England)*, 5(4), 368–373.
- Rambold, H., Churchland, A., Selig, Y., Jasmin, L., & Lisberger, S. G. (2002). Partial ablations of the flocculus and ventral paraflocculus in monkeys cause linked deficits in smooth pursuit eye movements and adaptive modification of the VOR. *Journal of Neurophysiology*, 87(2), 912–924.
- Randolph, M., & Semmes, J. (1974). Behavioral consequences of selective subtotal ablations in the postcentral gyrus of *Macaca mulatta*. *Brain Research*, 70(1), 55–70.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. doi:10.1038/4580.
- Rasch, B., Buchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, 315(5817), 1426–1429. doi:10.1126/science.1138581.
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8(4), 516–521.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724. doi:10.1038/nn.2331.
- Ray, S., Hsiao, S. S., Crone, N. E., Franaszczuk, P. J., & Niebur, E. (2008). Effect of stimulus intensity on the spike-local field potential relationship in the secondary somatosensory cortex. *Journal of Neuroscience*, 28(29), 7334–7343. doi:10.1523/JNEUROSCI.1588-08.2008.
- Recanzone, G. H., & Sutter, M. L. (2008). The biological basis of audition. *Annual Review of Psychology*, 59, 119–142. doi:10.1146/annurev.psych.59.103006.093544.
- Rechtschaffen, A., & Siegel, J. (2000). Sleep and dreaming. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (4th ed., pp. 936–947). New York: McGraw-Hill.
- Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience*, 3(4), 261–270.
- Reid, A. T., Krummack, A., Wanke, E., & Kotter, R. (2009). Optimization of cortical hierarchies with continuous scales and ranges. *NeuroImage*, 47(2), 611–617. doi:10.1016/j.neuroimage.2009.04.061.
- Reinders, A. A., Nijenhuis, E. R., Paans, A. M., Korf, J., Willemse, A. T., & den Boer, J. A. (2003). One brain, two selves. *NeuroImage*, 20(4), 2119–2125.
- Reiss, D., & Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: a case of cognitive convergence. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10), 5937–5942. doi:10.1073/pnas.101086398.
- Rempel-Clower, N. L., & Barbas, H. (2000). The laminar pattern of connections between prefrontal and anterior temporal cortices in the rhesus monkey is related to cortical structure and function. *Cerebral Cortex*, 10(9), 851–865.
- Remy, P., Zilbovicius, M., Cesaro, P., Amarenco, P., Degos, J. D., & Samson, Y. (1999). Primary somatosensory cortex activation is not altered in patients with ventroposterior thalamic lesions: a PET study. *Stroke*, 30(12), 2651–2658.
- Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965), 587–590. doi:10.1126/science.1179850.
- Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, 24(1), 19–29.

- Ribary, U., Ioannides, A. A., Singh, K. D., Hasson, R., Bolton, J. P., Lado, F., et al. (1991). Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 88(24), 11037–11041.
- Riddoch, G. (1917). Dissociation of visual perception due to occipital injuries, with especial reference to appreciation of movement. *Brain*, 40, 15–57.
- Riddoch, M. J., Humphreys, G. W., Akhtar, N., Allen, H., Bracewell, R. M., & Schofield, A. J. (2008). A tale of two agnosias: distinctions between form and integrative agnosia. *Cognitive Neuropsychology*, 25(1), 56–92. doi:10.1080/02643290701848901.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019–1025. doi:10.1038/14819.
- Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Current Opinion in Neurobiology*, 6(2), 228–236.
- Rockel, A. J., Hiorns, R. W., & Powell, T. P. (1980). The basic uniformity in structure of the neocortex. *Brain*, 103(2), 221–244.
- Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, 179(1), 3–20.
- Rockland, K. S., & Van Hoesen, G. W. (1994). Direct temporal–occipital feedback connections to striate cortex (V1) in the macaque monkey. *Cerebral Cortex*, 4(3), 300–313.
- Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397(6718), 430–433.
- Roe, A. W., Pallas, S. L., Hahm, J. O., & Sur, M. (1990). A map of visual space induced in primary auditory cortex. *Science*, 250(4982), 818–820.
- Roe, A. W., Parker, A. J., Born, R. T., & DeAngelis, G. C. (2007). Disparity channels in early vision. *Journal of Neuroscience*, 27(44), 11820–11831. doi:10.1523/JNEUROSCI.4164-07.2007.
- Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, 29, 203–227. doi:10.1146/annurev.neuro.29.051605.112939.
- Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience*, 7(9), 982–991.
- Rogan, M. T., Staubli, U. V., & LeDoux, J. E. (1997). Fear conditioning induces associative long-term potentiation in the amygdala. *Nature*, 390(6660), 604–607. doi:10.1038/37601.
- Rolls, E. T. (2012). Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. *Frontiers in Computational Neuroscience*, 6, 35. doi:10.3389/fncom.2012.00035.
- Romanelli, F., & Smith, K. M. (2009). Dextromethorphan abuse: clinical effects and management. *Journal of the American Pharmacists Association*, 49(2), e20–e25.
- Romei, V., Gross, J., & Thut, G. (2010). On the role of prestimulus alpha rhythms over occipito–parietal areas in visual input regulation: correlation or causation? *Journal of Neuroscience*, 30(25), 8692–8697. doi:10.1523/JNEUROSCI.0160-10.2010.
- Rorty, R. (1971). Mind–brain identity, privacy and categories. In D. Rosenthal (Ed.), *Materialism and the mind–body problem* (pp. 174–199). Englewood Cliffs, NJ: Prentice Hall.
- Rose, D. (1995). A portrait of the brain. In R. Gregory, J. Harris, P. Heard, & D. Rose (Eds.), *The artful eye* (pp. 28–51). Oxford: Oxford University Press.
- Rose, D. (2006). *Consciousness: philosophical, psychological and neural theories*. Oxford, UK: Oxford University Press.
- Rosenblatt, F. (1961). *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Washington, DC: Spartan Books.
- Rosse, R. B., Collins, J. P., Jr., Fay-McCarthy, M., Alim, T. N., Wyatt, R. J., & Deutsch, S. I. (1994). Phenomenologic comparison of the idiopathic psychosis of schizophrenia and drug-induced cocaine and phencyclidine psychoses: a retrospective study. *Clinical Neuropharmacology*, 17(4), 359–369.

- Roth, A., & Häusser, M. (2001). Compartmental models of rat cerebellar Purkinje cells based on simultaneous somatic and dendritic patch-clamp recordings. *Journal of Physiology*, 535(Pt 2), 445–472. doi:PHY_12145 [pii].
- Rotstein, H. G., Pervouchine, D. D., Acker, C. D., Gillies, M. J., White, J. A., Buhl, E. H., et al. (2005). Slow and fast inhibition and an H-current interact to create a theta rhythm in a model of CA1 interneuron network. *Journal of Neurophysiology*, 94(2), 1509–1518.
- Rozzi, S., Calzavara, R., Belmalih, A., Borra, E., Gregoriou, G. G., Matelli, M., et al. (2006). Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cerebral Cortex*, 16(10), 1389–1417. doi:10.1093/cercor/bhj076.
- Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition*, 31(6), 887–901.
- Rumelhart, D. E., & McClelland, J. L. (1985). Levels indeed! A response to Broadbent. *Journal of Experimental Psychology. General*, 114, 193–197.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition* (Vol. 1, pp. 316–362). Cambridge, MA: MIT Press.
- Russell, B. (1912). *The problems of philosophy*. London: Thornton Butterworth.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Sacks, O. (1973). *Awakenings*. London: Duckworth.
- Sacks, O. (1996). *An anthropologist on Mars*. New York: Picador, MacMillan.
- Sadleir, L. G., Scheffer, I. E., Smith, S., Connolly, M. B., & Farrell, K. (2009). Automatisms in absence seizures in children with idiopathic generalized epilepsy. *Archives of Neurology*, 66(6), 729–734. doi:10.1001/archneurol.2009.108.
- Sahraie, A., Hibbard, P. B., Trevethan, C. T., Ritchie, K. L., & Weiskrantz, L. (2010). Consciousness of the first order in blindsight. *Proceedings of the National Academy of Sciences of the United States of America*, 107(49), 21217–21222. doi:10.1073/pnas.1015652107.
- Sakai, K., & Jouvet, M. (1980). Brain stem PGO-on cells projecting directly to the cat dorsal lateral geniculate nucleus. *Brain Research*, 194(2), 500–505.
- Salenius, S., Schnitzler, A., Salmelin, R., Jousmaki, V., & Hari, R. (1997). Modulation of human cortical rolandic rhythms during natural sensorimotor tasks. *NeuroImage*, 5(3), 221–228. doi:10.1006/nimg.1997.0261.
- Salinas, E., & Sejnowski, T. J. (2000). Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *Journal of Neuroscience*, 20(16), 6193–6209.
- Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346(6280), 174–177. doi:10.1038/346174a0.
- Sanchez-Vives, M. V., & McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neuroscience*, 3(10), 1027–1034. doi:10.1038/79848.
- Satterlie, R. A. (2011). Do jellyfish have central nervous systems? *Journal of Experimental Biology*, 214(Pt 8), 1215–1223. doi:10.1242/jeb.043687.
- Saul, A. B., Carras, P. L., & Humphrey, A. L. (2005). Temporal properties of inputs to direction-selective neurons in monkey V1. *Journal of Neurophysiology*, 94(1), 282–294.
- Savander, V., Miettinen, R., Ledoux, J. E., & Pitkänen, A. (1997). Lateral nucleus of the rat amygdala is reciprocally connected with basal and accessory basal nuclei: a light and electron microscopic study. *Neuroscience*, 77(3), 767–781.
- Saxena, A., Driemeyer, J., & Ng, A. Y. (2008). Robotic grasping of novel objects using vision. *International Journal of Robotics Research*, 27(2), 157–173. doi:10.1177/0278364907087172.
- Scannell, J. W., & Young, M. P. (1993). The connectional organization of neural systems in the cat cerebral cortex. *Current Biology*, 3(4), 191–200. doi:0960–9822(93)90331-H [pii].
- Schenk, T., & Zihl, J. (1997). Visual motion perception after brain damage: I. Deficits in global motion perception. *Neuropsychologia*, 35(9), 1289–1297.

- Schiff, N. D., Giacino, J. T., Kalmar, K., Victor, J. D., Baker, K., Gerber, M., et al. (2007). Behavioural improvements with thalamic stimulation after severe traumatic brain injury. *Nature*, 448(7153), 600–603. doi:10.1038/nature06041.
- Schiff, N. D., Nauvel, T., & Victor, J. D. (2014). Large-scale brain dynamics in disorders of consciousness. *Current Opinion in Neurobiology*, 25C, 7–14. doi:10.1016/j.conb.2013.10.007.
- Schiff, N. D., & Plum, F. (2000). The role of arousal and “gating” systems in the neurology of impaired consciousness. *Journal of Clinical Neurophysiology*, 17(5), 438–452.
- Schiff, N. D., Ribary, U., Moreno, D. R., Beattie, B., Kronberg, E., Blasberg, R., et al. (2002). Residual cerebral activity and behavioural fragments can remain in the persistently vegetative brain. *Brain*, 125(Pt 6), 1210–1234.
- Schiller, P. H., & Tehovnik, E. J. (2005). Neural mechanisms underlying target selection with saccadic eye movements. *Progress in Brain Research*, 149, 157–171.
- Schlaggar, B. L., & O’Leary, D. D. (1991). Potential of visual cortex to develop an array of functional units unique to somatosensory cortex. *Science*, 252(5012), 1556–1560.
- Schmid, M. C., Mrowka, S. W., Turchi, J., Saunders, R. C., Wilke, M., Peters, A. J., et al. (2010). Blindsight depends on the lateral geniculate nucleus. *Nature*, 466(7304), 373–377. doi:10.1038/nature09179.
- Schnitzler, A., & Ploner, M. (2000). Neurophysiology and functional neuroanatomy of pain perception. *Journal of Clinical Neurophysiology*, 17(6), 592–603.
- Schoenbaum, G., Roesch, M. R., Stalnaker, T. A., & Takahashi, Y. K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews Neuroscience*, 10(12), 885–892. doi:10.1038/nrn2753.
- Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W., & Lamme, V. A. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision*, 9(4), 29, 1–15. doi:10.1167/9.4.29
- Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Research. Cognitive Brain Research*, 14(1), 187–198.
- Schroeder, C. E., & Lakatos, P. (2008). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1), 9–18.
- Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology*, 85(3), 1322–1327.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599.
- Schwarz, H. D., Esteky, H., & Jones, E. G. (1992). Corticocortical connections of cat primary somatosensory cortex. *Experimental Brain Research*, 91(3), 425–434.
- Schwartz, S., & Maquet, P. (2002). Sleep imaging and the neuro-psychological assessment of dreams. *Trends in Cognitive Sciences*, 6(1), 23–30.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20(1), 11–21.
- Seager, W. (1995). Consciousness, information and panpsychism. *Journal of Consciousness Studies*, 2, 272–288.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417–457.
- Searle, J. R. (1983). *Intentionality: an essay in the philosophy of the mind*. Cambridge: Cambridge University Press.
- Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience*, 23, 557–578.
- Searle, J. R. (2004). *Mind*. Oxford: Oxford University Press.
- Sebe, J. Y., & Berger, A. J. (2008). Inspiratory-phase short time scale synchrony in the brainstem slice is generated downstream of the pre-Botzinger complex. *Neuroscience*, 153(4), 1390–1401. doi:10.1016/j.neuroscience.2008.02.034.

- Seo, D., Carmena, J. M., Rabaey, J. M., Alon, E., & Maharbiz, M. M. (2013). Neural dust: An ultrasonic, low power solution for chronic brain-machine interfaces. *arXiv*, 1307.2196v1301.
- Sereno, M. I., & Huang, R. S. (2006). A human parietal face area contains aligned head-centered visual and tactile maps. *Nature Neuroscience*, 9(10), 1337–1343.
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, 8(10), 1391–1400. doi:10.1038/nrn1549.
- Seth, A. K. (2008). Causal networks in simulated neural systems. *Cognitive Neurodynamics*, 2(1), 49–64. doi:10.1007/s11571-007-9031-z.
- Seth, A. K., Izhikevich, E., Rekke, G. N., & Edelman, G. M. (2006). Theories and measures of consciousness: an extended framework. *Proceedings of the National Academy of Sciences of the United States of America*, 103(28), 10799–10804. doi:10.1073/pnas.0604347103.
- Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron*, 24, 67–77.
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4), 1916–1936.
- Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16), 10917–10922. doi:10.1073/pnas.152694799.
- Shambes, G. M., Gibson, J. M., & Welker, W. (1978). Fractured somatotopy in granule cell tactile areas of rat cerebellar hemispheres revealed by micromapping. *Brain, Behavior and Evolution*, 15(2), 94–140.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers*, 37, 10–21.
- Scheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences of the United States of America*, 94(7), 3408–3413.
- Shepherd, G. (1983). *Neurobiology*. Oxford: Oxford University Press.
- Shepherd, G. M. (Ed.) (1990). *The synaptic organization of the brain* (3rd ed.). Oxford: Oxford University Press.
- Shepherd, G. M. (2007). Perspectives on olfactory processing, conscious perception, and orbitofrontal cortex. *Annals of the New York Academy of Sciences*, 1121, 87–101. doi:10.1196/annals.1401.032.
- Shepherd, G. M. (2009). Intracortical cartography in an agranular area. *Frontiers in Neuroscience*, 3, 337–343.
- Sherman, S. M. (2006). What is the function of the thalamus? In J. L. Van Hemmen & T. J. Sejnowski (Eds.), *23 problems in systems neuroscience* (pp. 65–82). Oxford: Oxford University Press.
- Sherrington, C. S. (1906). *The integrative action of the nervous system*. New Haven, CT: Yale University Press.
- Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, 11(4), 505–509.
- Shindy, W. W., Posley, K. A., & Fuster, J. M. (1994). Reversible deficit in haptic delay tasks from cooling prefrontal cortex. *Cerebral Cortex*, 4(4), 443–450.
- Shoemaker, S. (1990). Qualities and qualia: what's in the mind? *Philosophy and Phenomenological Research*, 50(Suppl.), 109–131.
- Shoemaker, S. (1994). Phenomenal character. *Noûs*, 28(1), 21–38. doi:10.2307/2215918.
- Shuler, M. G., & Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science*, 311(5767), 1606–1609. doi:10.1126/science.1123513.
- Siapas, A. G., & Wilson, M. A. (1998). Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron*, 21(5), 1123–1128.
- Siegel, J. M. (2001). The REM sleep–memory consolidation hypothesis. *Science*, 294(5544), 1058–1063. doi:10.1126/science.1063049.

- Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., & Engel, A. K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. *Neuron*, 60(4), 709–719. doi:10.1016/j.neuron.2008.09.010.
- Siegel, M., Warden, M. R., & Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), 21341–21346. doi:10.1073/pnas.0908193106.
- Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nature Neuroscience*, 8(2), 143–144. doi:10.1038/nn1379.
- Singh-Curry, V., & Husain, M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia*, 47(6), 1434–1448. doi:10.1016/j.neuropsychologia.2008.11.033.
- Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271(5257), 1870–1873.
- Sloan, T. B. (1998). Anesthetic effects on electrophysiologic recordings. *Journal of Clinical Neurophysiology*, 15(3), 217–226.
- Small, D. M. (2010). Taste representation in the human insula. *Brain Structure & Function*, 214(5–6), 551–561. doi:10.1007/s00429-010-0266-9.
- Small, D. M., Gerber, J. C., Mak, Y. E., & Hummel, T. (2005). Differential neural responses evoked by orthonasal versus retronasal odorant perception in humans. *Neuron*, 47(4), 593–605.
- Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, 68, 141–156.
- Smith, M. A., Jia, X., Zandvakili, A., & Kohn, A. (2013). Laminar dependence of neuronal correlations in visual cortex. *Journal of Neurophysiology*, 109(4), 940–947. doi:10.1152/jn.00846.2012.
- Smith, P. H., Manning, K. A., & Uhlrich, D. J. (2010). Evaluation of inputs to rat primary auditory cortex from the suprageniculate nucleus and extrastriate visual cortex. *Journal of Comparative Neurology*, 518(18), 3679–3700. doi:10.1002/cne.22411.
- Smith, P. H., & Populin, L. C. (2001). Fundamental differences between the thalamocortical recipient layers of the cat auditory and visual cortices. *Journal of Comparative Neurology*, 436(4), 508–519.
- Smolensky, P. (1987). Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review*, 1, 95–109.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1–74.
- Snyder, L. H., Grieve, K. L., Brotchie, P., & Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature*, 394(6696), 887–891. doi:10.1038/29777.
- Solomon, S. G., & Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience*, 8(4), 276–286. doi:10.1038/nrn2094.
- Sommer, I. E., Selten, J. P., Diederen, K. M., & Blom, J. D. (2010). Dissecting auditory verbal hallucinations into two components: audibility (Gedankenlautwerden) and alienation (thought insertion). *Psychopathology*, 43(2), 137–140. doi:10.1159/000277003.
- Sommer, M. A., & Wurtz, R. H. (2008). Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience*, 31, 317–338. doi:10.1146/annurev.neuro.31.060407.125627.
- Sompolinsky, H., & Kanter, I. I. (1986). Temporal association in asymmetric neural networks. *Physical Review Letters*, 57(22), 2861–2864.
- Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *NeuroImage*, 35(2), 836–852. doi:10.1016/j.neuroimage.2006.09.051.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74, 1–29.
- Spinoza, B. de (1677). *Ethica* (Ethica Ordine Geometrico Demonstrata; 1951 translation by R. H. M. Elwes). Auckland, NZ: Floating Press.
- Sporns, O., Honey, C. J., & Kotter, R. (2007). Identification and classification of hubs in brain networks. *PLoS ONE*, 2(10), e1049. doi:10.1371/journal.pone.0001049.

- Squire, L. R., Shimamura, A. P., & Amaral, D. G. (1989). Memory and the hippocampus. In J. H. Byrne & W. O. Berry (Eds.), *Neural models of plasticity: experimental and theoretical approaches* (p. 227). New York: Academic Press.
- Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279–306. doi:10.1146/annurev.neuro.27.070203.144130.
- Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 216(1205), 427–459.
- Stark, E., Drori, R., & Abeles, M. (2006). Partial cross-correlation analysis resolves ambiguity in the encoding of multiple movement features. *Journal of Neurophysiology*, 95(3), 1966–1975.
- Stefanacci, L., & Amaral, D. G. (2002). Some observations on cortical inputs to the macaque monkey amygdala: an anterograde tracing study. *Journal of Comparative Neurology*, 451(4), 301–323. doi:10.1002/cne.10339.
- Stefanacci, L., Buffalo, E. A., Schmolck, H., & Squire, L. R. (2000). Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neuropsychological profile of patient E. P. *Journal of Neuroscience*, 20(18), 7024–7036.
- Stein, B. E., & Gaither, N. S. (1981). Sensory representation in reptilian optic tectum: some comparisons with mammals. *Journal of Comparative Neurology*, 202(1), 69–87. doi:10.1002/cne.902020107.
- Stein, B. E., Jiang, W., Wallace, M. T., & Stanford, T. R. (2001). Nonvisual influences on visual-information processing in the superior colliculus. *Progress in Brain Research*, 134, 143–156.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- Steinbeck, J. (1962). *Travels with Charley*. New York: Penguin Books.
- Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B., & Grillner, S. (2011). Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Current Biology*, 21(13), 1081–1091. doi:10.1016/j.cub.2011.05.001.
- Stepniewska, I., Preuss, T. M., & Kaas, J. H. (1993). Architectonics, somatotopic organization, and ipsilateral cortical connections of the primary motor area (M1) of owl monkeys. *Journal of Comparative Neurology*, 330(2), 238–271.
- Steriade, M. (1996). Arousal: revisiting the reticular activating system. *Science*, 272(5259), 225–226.
- Steriade, M. (2004). Acetylcholine systems and rhythmic activities during the waking–sleep cycle. *Progress in Brain Research*, 145, 179–196.
- Steriade, M., Contreras, D., Amzica, F., & Timofeev, I. (1996). Synchronization of fast (30–40 Hz) spontaneous oscillations in intrathalamic and thalamocortical networks. *Journal of Neuroscience*, 16(8), 2788–2808.
- Steriade, M., Contreras, D., Curro Dossi, R., & Nunez, A. (1993). The slow (< 1 Hz) oscillation in reticular thalamic and thalamocortical neurons: scenario of sleep rhythm generation in interacting thalamic and neocortical networks. *Journal of Neuroscience*, 13(8), 3284–3299.
- Steriade, M., McCormick, D. A., & Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262(5134), 679–685.
- Steriade, M., Nunez, A., & Amzica, F. (1993). A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *Journal of Neuroscience*, 13(8), 3252–3265.
- Steriade, M., Pare, D., Datta, S., Oakson, G., & Curro Dossi, R. (1990). Different cellular types in mesopontine cholinergic nuclei related to ponto–geniculo–occipital waves. *Journal of Neuroscience*, 10(8), 2560–2579.
- Steriade, M., Timofeev, I., & Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *Journal of Neurophysiology*, 85(5), 1969–1985.
- Stern, E. A., Jaeger, D., & Wilson, C. J. (1998). Membrane potential synchrony of simultaneously recorded striatal spiny neurons in vivo. *Nature*, 394(6692), 475–478. doi:10.1038/28848.
- Stewart, R. M. (1925). On a case of organic spinal hemianesthesia showing the persistence of a peculiar form of sensation and the occurrence of phenomena of alloesthesia and heteraesthesia. *Journal of Neurology and Psychopathology*, 5(20), 289–317.
- Stoerig, P., & Cowey, A. (1997). Blindsight in man and monkey. *Brain*, 120(Pt 3), 535–559.

- Stolzenberg, D. S., & Numan, M. (2011). Hypothalamic interaction with the mesolimbic DA system in the control of the maternal and sexual behaviors in rats. *Neuroscience and Biobehavioral Reviews*, 35(3), 826–847.
- Strawson, G. (2006). *Consciousness and its place in nature*. Exeter, UK: Imprint Academic.
- Striedter, G. F. (2006). Precis of principles of brain evolution. *Behavioral and Brain Sciences*, 29(1), 1–12, discussion 12–36. doi:10.1017/S0140525X06009010.
- Stuart, G., & Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, 367(6458), 69–72.
- Stuart, G., Spruston, N., Sakmann, B., & Häusser, M. (1997). Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends in Neurosciences*, 20(3), 125–131. doi:S0166-2236(96)10075-8 [pii].
- Sufczynski, P., Kalitzin, S., Pfurtscheller, G., & Lopes da Silva, F. H. (2001). Computational model of thalamo-cortical networks: dynamical control of alpha rhythms in relation to focal attention. *International Journal of Psychophysiology*, 43(1), 25–40.
- Suga, N. (1989). Principles of auditory information-processing derived from neuroethology. *Journal of Experimental Biology*, 146, 277–286.
- Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature*, 401(6750), 269–272.
- Sugita, Y. (2004). Experience in early infancy is indispensable for color perception. *Current Biology*, 14(14), 1267–1271. doi:10.1016/j.cub.2004.07.020.
- Sun, R., & Bookman, L. (1994). *Computational architectures integrating neural and symbolic processes*. Needham, MA: Kluwer Academic.
- Supèr, H., Spekreijse, H., & Lamme, V. A. (2001a). A neural correlate of working memory in the monkey primary visual cortex. *Science*, 293(5527), 120–124. doi:10.1126/science.1060496.
- Supèr, H., Spekreijse, H., & Lamme, V. A. (2001b). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience*, 4(3), 304–310. doi:10.1038/85170.
- Supp, G. G., Siegel, M., Hipp, J. F., & Engel, A. K. (2011). Cortical hypersynchrony predicts breakdown of sensory processing during loss of consciousness. *Current Biology*, 21(23), 1988–1993. doi:10.1016/j.cub.2011.10.017.
- Sutcliffe, J. G., & de Lecea, L. (2002). The hypocretins: setting the arousal threshold. *Nature Reviews Neuroscience*, 3(5), 339–349.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.
- Suzuki, H., Uchiyama, M., Tagaya, H., Ozaki, A., Kuriyama, K., Aritake, S., et al. (2004). Dreaming during non-rapid eye movement sleep in the absence of prior rapid eye movement sleep. *Sleep*, 27(8), 1486–1490.
- Sylvester, C. M., Krout, K. E., & Loewy, A. D. (2002). Suprachiasmatic nucleus projection to the medial prefrontal cortex: a viral transneuronal tracing study. *Neuroscience*, 114(4), 1071–1080. doi:S0306452202003615 [pii].
- Tahvildari, B., Wolfel, M., Duque, A., & McCormick, D. A. (2012). Selective functional interactions between excitatory and inhibitory cortical neurons and differential contribution to persistent activity of the slow oscillation. *Journal of Neuroscience*, 32(35), 12165–12179. doi:10.1523/JNEUROSCI.1181-12.2012.
- Tallon-Baudry, C., Bertrand, O., & Fischer, C. (2001). Oscillatory synchrony between human extrastriate areas during visual short-term memory maintenance. *Journal of Neuroscience*, 21(20), RC177.
- Tallon-Baudry, C., Bertrand, O., Peronnet, F., & Pernier, J. (1998). Induced gamma-band activity during the delay of a visual short-term memory task in humans. *Journal of Neuroscience*, 18(11), 4244–4254.
- Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, 30(43), 14356–14360.
- Tamura, H., & Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex*, 11(5), 384–399.
- Tanabe, T., Iino, M., & Takagi, S. F. (1975). Discrimination of odors in olfactory bulb, pyriform–amygdaloid areas, and orbitofrontal cortex of the monkey. *Journal of Neurophysiology*, 38(5), 1284–1296.
- Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*, 13(1), 90–99.

- Tang, J., & Suga, N. (2009). Corticocortical interactions between and within three cortical auditory areas specialized for time-domain signal processing. *Journal of Neuroscience*, 29(22), 7230–7237. doi:10.1523/JNEUROSCI.0373-09.2009.
- Tanila, H., Shapiro, M. L., & Eichenbaum, H. (1997). Discordance of spatial representation in ensembles of hippocampal place cells. *Hippocampus*, 7(6), 613–623. doi:10.1002/(SICI)1098-1063(1997)7:6<613::AID-HIPO4>3.0.CO;2-F.
- Tatsuno, M., Lipa, P., & McNaughton, B. L. (2006). Methodological considerations on the use of template matching to study long-lasting memory trace replay. *Journal of Neuroscience*, 26(42), 10727–10742. doi:10.1523/JNEUROSCI.3317-06.2006.
- Taverna, S., van Dongen, Y. C., Groenewegen, H. J., & Pennartz, C. M. (2004). Direct physiological evidence for synaptic connectivity between medium-sized spiny neurons in rat nucleus accumbens in situ. *Journal of Neurophysiology*, 91(3), 1111–1121. doi:10.1152/jn.00892.2003.
- Tesauro, G., & Sejnowski, T. J. (1989). A parallel network that learns to play backgammon. *Artificial Intelligence Journal*, 39, 357–390.
- Teuber, H. L. (1968). Alteration of perception and memory in man. In L. Weiskrantz (Ed.), *Analysis of behavioral change* (pp. 268–375). New York: Harper & Row.
- Theyel, B. B., Llano, D. A., & Sherman, S. M. (2010). The corticothalamicocortical circuit drives higher-order cortex in the mouse. *Nature Neuroscience*, 13(1), 84–88. doi:10.1038/nn.2449.
- Thiebaut de Schotten, M., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G., et al. (2011). A lateralized brain network for visuospatial attention. *Nature Neuroscience*, 14(10), 1245–1246. doi:10.1038/nn.2905.
- Thier, P., & Möck, M. (2006). The oculomotor role of the pontine nuclei and the nucleus reticularis tegmenti pontis. *Progress in Brain Research*, 151, 293–320.
- Thompson, E., & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Sciences*, 5(10), 418–425.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522. doi:10.1038/381520a0.
- Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, 292(5515), 290–293. doi:10.1126/science.1058911.
- Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, 29(3), 757–767.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189–208.
- Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21(4), 753–759.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(1), 42.
- Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, 282(5395), 1846–1851.
- Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of the United States of America*, 91(11), 5033–5037.
- Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E., et al. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420(6917), 788–794.
- Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. (2001). Neurobiological foundations for the theory of harmony in Western tonal music. *Annals of the New York Academy of Sciences*, 930, 92–116.
- Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 353(1373), 1295–1306.
- Treisman, A. (1999). Solutions to the binding problem: progress through controversy and convergence. *Neuron*, 24, 105–110.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.

- Treves, A. (2005). Frontal latching networks: a possible neural basis for infinite recursion. *Cognitive Neuropsychology*, 22(3), 276–291. doi:10.1080/02643290442000329.
- Trifa, V. M., Koene, A., Jan Moren, J., & Cheng, G. (2007). Real-time acoustic source localization in noisy environments for human–robot multimodal interaction. Paper presented at the 16th IEEE International Conference on Robot and Human Interactive Communication, Jeju, Korea.
- Truccolo, W., Hochberg, L. R., & Donoghue, J. P. (2010). Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nature Neuroscience*, 13(1), 105–111. doi:10.1038/nrn2455.
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., et al. (2007). Schemas and memory consolidation. *Science*, 316(5821), 76–82. doi:10.1126/science.1135935.
- Ts'o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience*, 6(4), 1160–1170.
- Tsodyks, M. (1999). Attractor neural network models of spatial maps in hippocampus. *Hippocampus*, 9(4), 481–489. doi:10.1002/(SICI)1098-1063(1999)9:4<481::AID-HIPO14>3.0.CO;2-S.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.
- Tunstall, M. J., Oorschot, D. E., Kean, A., & Wickens, J. R. (2002). Inhibitory interactions between spiny projection neurons in the rat striatum. *Journal of Neurophysiology*, 88(3), 1263–1269.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 236, 433–460.
- Turton, A. J., & Butler, S. R. (2001). Referred sensations following stroke. *Neurocase*, 7(5), 397–405. doi:10.1076/neur.7.5.397.16251.
- Tye, M. (1994). Qualia, content, and the inverted-spectrum. *Nous*, 28(2), 159–183. doi:10.2307/2216047.
- Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: MIT Press.
- Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.
- Vahle-Hinz, C., Detsch, O., Siemers, M., & Kochs, E. (2007). Contributions of GABAergic and glutamatergic mechanisms to isoflurane-induced suppression of thalamic somatosensory information transfer. *Experimental Brain Research*, 176(1), 159–172. doi:10.1007/s00221-006-0604-6.
- Valenstein, E. S. (1986). *Great and desperate cures: the rise and decline of psychosurgery and other radical treatments for mental illness*. New York: Basic Books.
- Vallar, G. (1998). Spatial hemineglect in humans. *Trends in Cognitive Sciences*, 2(3), 87–97.
- Vallar, G., Bottini, G., Sterzi, R., Passerini, D., & Rusconi, M. L. (1991). Hemianesthesia, sensory neglect, and defective access to conscious experience. *Neurology*, 41(5), 650–652.
- Vallar, G., Papagno, C., & Baddeley, A. D. (1991). Long-term recency effects and phonological short-term memory. A neuropsychological case study. *Cortex*, 27(2), 323–326.
- Vallar, G., & Ronchi, R. (2009). Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Experimental Brain Research*, 192(3), 533–551. doi:10.1007/s00221-008-1562-y.
- van den Pol, A. N., & Tsujimoto, K. L. (1985). Neurotransmitters of the hypothalamic suprachiasmatic nucleus: immunocytochemical analysis of 25 neuronal antigens. *Neuroscience*, 15(4), 1049–1086.
- van der Meer, M. A., & Redish, A. D. (2009). Low and high gamma oscillations in rat ventral striatum have distinct relationships to behavior, reward, and spiking activity on a learned spatial decision task. *Frontiers in Integrative Neuroscience*, 3, 9. doi:10.3389/neuro.07.009.2009.
- Van der Werf, Y. D., Witter, M. P., & Groenewegen, H. J. (2002). The intralaminar and midline nuclei of the thalamus: anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Research. Brain Research Reviews*, 39(2–3), 107–140. doi:S0165017302001819 [pii].
- van Dijk, H., Schoffelen, J. M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *Journal of Neuroscience*, 28(8), 1816–1823. doi:10.1523/JNEUROSCI.1853-07.2008.
- van Dongen, Y. C., Deniau, J. M., Pennartz, C. M., Galis-de Graaf, Y., Voorn, P., Thierry, A. M., et al. (2005). Anatomical evidence for direct connections between the shell and core subregions of the rat nucleus accumbens. *Neuroscience*, 136(4), 1049–1071. doi:10.1016/j.neuroscience.2005.08.050.

- van Duuren, E., Lankelma, J., & Pennartz, C. M. (2008). Population coding of reward magnitude in the orbitofrontal cortex of the rat. *Journal of Neuroscience*, 28(34), 8590–8603.
- van Duuren, E., van der Plasse, G., Lankelma, J., Joosten, R. N., Feenstra, M. G., & Pennartz, C. M. (2009). Single-cell and population coding of expected reward probability in the orbitofrontal cortex of the rat. *Journal of Neuroscience*, 29(28), 8965–8976. doi:10.1523/JNEUROSCI.0005-09.2009.
- van Ee, R., van Dam, L. C., & Brouwer, G. J. (2005). Voluntary control and the dynamics of perceptual bistability. *Vision Research*, 45(1), 41–55. doi:10.1016/j.visres.2004.07.030.
- Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science*, 255(5043), 419–423.
- Van Gulick, R. (1989). What difference does consciousness make? *Philosophical Topics*, 17, 211–230.
- Van Gulick, R. (2001). Reduction, emergence and other recent options on the mind/body problem: a philosophic overview. *Journal of Consciousness Studies*, 8, 1–34.
- van Kesteren, M. T., Fernandez, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal–neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), 7550–7555. doi:10.1073/pnas.0914892107.
- van Luterveld, R., Sommer, I. E., & Ford, J. M. (2011). The neurophysiology of auditory hallucinations—a historical and contemporary review. *Frontiers in Psychiatry*, 2, 28. doi:10.3389/fpsyg.2011.00028.
- van Wingerden, M., Vinck, M., Lankelma, J. V., & Pennartz, C. M. (2010). Learning-associated gamma-band phase-locking of action-outcome selective neurons in orbitofrontal cortex. *Journal of Neuroscience*, 30(30), 10025–10038. doi:10.1523/JNEUROSCI.0222-10.2010.
- Vanderwolf, C. H. (1969). Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalography and Clinical Neurophysiology*, 26(4), 407–418.
- Vanderwolf, C. H., & Stewart, D. J. (1988). Thalamic control of neocortical activation: a critical re-evaluation. *Brain Research Bulletin*, 20(4), 529–538.
- VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5), 207–213.
- VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, 42(23), 2593–2615.
- Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., & Schwartz, A. B. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature*, 453(7198), 1098–1101. doi:10.1038/nature06996.
- Velly, L. J., Rey, M. F., Bruder, N. J., Gouvitsois, F. A., Witjas, T., Regis, J. M., et al. (2007). Differential dynamic of action on cortical and subcortical structures of anesthetic agents during induction of anesthesia. *Anesthesiology*, 107(2), 202–212. doi:10.1097/01.anes.0000270734.99298.b4.
- Veselis, R. A., Feshchenko, V. A., Reinsel, R. A., Dnistrian, A. M., Beattie, B., & Akhurst, T. J. (2004). Thiopental and propofol affect different regions of the brain at similar pharmacologic effects. *Anesthesia and Analgesia*, 99(2), 399–408. doi:10.1213/01.ANE.0000131971.92180.DF
- Vezoli, J., Falchier, A., Jouve, B., Knoblauch, K., Young, M., & Kennedy, H. (2004). Quantitative analysis of connectivity in the visual cortex: extracting function from structure. *Neuroscientist*, 10(5), 476–482. doi:10.1177/1073858404268477.
- Vilensky, J. A., Gilman, S., & McCall, S. (2010). Does the historical literature on encephalitis lethargica support a simple (direct) relationship with postencephalitic Parkinsonism? *Movement Disorders*, 25(9), 1124–1130. doi:10.1002/mds.22991.
- Villablanca, J., & Salinas-Zeballos, M. E. (1972). Sleep–wakefulness, EEG and behavioral studies of chronic cats without the thalamus: the “athalamic” cat. *Archives Italiennes de Biologie*, 110(3), 383–411.
- Vinck, M., Lima, B., Womelsdorf, T., Oostenveld, R., Singer, W., Neuenschwander, S., et al. (2010). Gamma-phase shifting in awake monkey visual cortex. *Journal of Neuroscience*, 30(4), 1250–1257. doi:10.1523/JNEUROSCI.1623-09.2010.
- Vinogradov, S., Luks, T. L., Schulman, B. J., & Simpson, G. V. (2008). Deficit in a neural correlate of reality monitoring in schizophrenia patients. *Cerebral Cortex*, 18(11), 2532–2539. doi:10.1093/cercor/bhn028.

- Vogt, B. A., & Laureys, S. (2005). Posterior cingulate, precuneal and retrosplenial cortices: cytology and components of the neural network correlates of consciousness. *Progress in Brain Research*, 150, 205–217. doi:10.1016/S0079-6123(05)50015-3.
- von Economo, K. (1917). Encephalitis lethargica. *Wiener Klinische Wochenschrift*, 30, 581–585.
- von der Malsburg, C. (1981). The correlation theory of brain function. MPI Biophysical Chemistry, Internal Report 81–2). In E. Domany, J. L. Van Hemmen, & K. Schulten (Eds.), *Models of neural networks II* (1994 ed., pp. 95–119). Berlin: Springer.
- von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5(4), 520–526.
- von der Malsburg, C. (1999). The what and why of binding: the modeler's perspective. *Neuron*, 24, 95–104.
- Von Hofsten, C. (1982). Eye-hand coordination in the newborn. *Developmental Psychology*, 18, 450–461.
- von Linstow Roloff, E., Harbaran, D., Micheau, J., Platt, B., & Riedel, G. (2007). Dissociation of cholinergic function in spatial and procedural learning in rats. *Neuroscience*, 146(3), 875–889. doi:10.1016/j.neuroscience.2007.02.038.
- von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, 404(6780), 871–876.
- von Senden, M. (1932). *Raum- und Gestaltauffassung bei Operierten Blindgeborenen* (translated as *Space and sight* by P. Heath, 1960, London: Methuen). Leipzig: Barth.
- Voorneveld, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. M. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences*, 27(8), 468–474. doi:10.1016/j.tins.2004.06.006.
- Wada, Y., & Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *Journal of Neurology, Neurosurgery, and Psychiatry*, 71(2), 254–257.
- Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, 427(6972), 352–355. doi:10.1038/nature02223.
- Walton, M. E., Croxson, P. L., Behrens, T. E., Kennerley, S. W., & Rushworth, M. F. (2007). Adaptive decision making and value in the anterior cingulate cortex. *NeuroImage*, 36(Suppl 2), T142–T154.
- Wamsley, E. J., Tucker, M., Payne, J. D., Benavides, J. A., & Stickgold, R. (2010). Dreaming of a learning task is associated with enhanced sleep-dependent memory consolidation. *Current Biology*, 20(9), 850–855. doi:10.1016/j.cub.2010.03.027.
- Wang, X. J., & Buzsaki, G. (1996). Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *Journal of Neuroscience*, 16(20), 6402–6413.
- Ward, R., Danziger, S., Owen, V., & Rafal, R. (2002). Deficits in spatial coding and feature binding following damage to spatiotopic maps in the human pulvinar. *Nature Neuroscience*, 5(2), 99–100. doi:10.1038/nn794.
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107(Pt 3), 829–854.
- Watanabe, M., Cheng, K., Murayama, Y., Ueno, K., Asamizuya, T., Tanaka, K., et al. (2011). Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science*, 334(6057), 829–831. doi:10.1126/science.1203161.
- Watkins, J. C., & Evans, R. H. (1981). Excitatory amino acid transmitters. *Annual Review of Pharmacology and Toxicology*, 21, 165–204.
- Watson, R. T., Heilman, K. M., Cauthen, J. C., & King, F. A. (1973). Neglect after cingulectomy. *Neurology*, 23(9), 1003–1007.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 440–442. doi:10.1038/30918.
- Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex*, 4(5), 470–483.
- Weir, C. J., Bradford, A. P., & Lees, K. R. (2003). The prognostic value of the components of the Glasgow Coma Scale following acute stroke. *QJM*, 96(1), 67–74.

- Weiskrantz, L. (1995). The problem of animal consciousness in relation to neuropsychology. *Behavioural Brain Research*, 71(1–2), 171–175.
- Weiskrantz, L. (1997). *Consciousness lost and found: a neuropsychological exploration*. Oxford: Oxford University Press.
- Weiskrantz, L., Barbur, J. L., & Sahraie, A. (1995). Parameters affecting conscious versus unconscious visual discrimination with damage to the visual cortex (V1). *Proceedings of the National Academy of Sciences of the United States of America*, 92(13), 6122–6126.
- Werbos, P. (1974). *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University.
- Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, 13(1), 1–7.
- Wheeler, J. A. (1990). Information, physics, quantum: the search for links. In W. Zurek (Ed.), *Complexity, entropy, and the physics of information* (pp. 309–336). Redwood City, CA: Addison-Wesley.
- Whishaw, I. Q., & Brooks, B. L. (1999). Calibrating space: exploration is important for allothetic and idiothetic navigation. *Hippocampus*, 9(6), 659–667.
- Widrow, B., Gupta, N. K., & Maitra, S. (1973). Punish/reward: learning with a critic in adaptive threshold systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 3, 455–465.
- Wierzyński, C. M., Lubenov, E. V., Gu, M., & Siapas, A. G. (2009). State-dependent spike-timing relationships between hippocampal and prefrontal circuits during sleep. *Neuron*, 61(4), 587–596. doi:10.1016/j.neuron.2009.01.011.
- Williams-Gray, C. H., Foltyne, T., Brayne, C. E., Robbins, T. W., & Barker, R. A. (2007). Evolution of cognitive dysfunction in an incident Parkinson's disease cohort. *Brain*, 130(Pt 7), 1787–1798. doi:10.1093/brain/awm111.
- Willis, W. D., Jr. (2007). The somatosensory system, with emphasis on structures important for pain. *Brain Research. Brain Research Reviews*, 55(2), 297–313.
- Willis, W. D., & Westlund, K. N. (1997). Neuroanatomy of the pain system and of the pathways that modulate pain. *Journal of Clinical Neurophysiology*, 14(1), 2–31.
- Willshaw, D. J. (1989). Holography, associative memory and inductive generalization. In G. Hinton & J. Anderson (Eds.), *Parallel models of associative memory* (pp. 103–124). Hillsdale, NJ: Erlbaum.
- Wilson, D. A., & Stevenson, R. J. (2003). Olfactory perceptual learning: the critical role of memory in odor discrimination. *Neuroscience and Biobehavioral Reviews*, 27(4), 307–328.
- Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, 261(5124), 1055–1058.
- Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172), 676–679.
- Wilson, R. S., Segawa, E., Hizel, L. P., Boyle, P. A., & Bennett, D. A. (2012). Terminal dedifferentiation of cognitive abilities. *Neurology*, 78(15), 1116–1122. doi:10.1212/WNL.0b013e31824f7ff2.
- Wimsatt, W. C. (1994). The ontology of complex systems: levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy*, 20, 207–274.
- Wise, S. P. (2008). Forward frontal fields: phylogeny and fundamental function. *Trends in Neurosciences*, 31(12), 599–608. doi:10.1016/j.tins.2008.08.008.
- Wise, S. P., & Jones, E. G. (1977). Cells of origin and terminal distribution of descending projections of the rat somatic sensory cortex. *Journal of Comparative Neurology*, 175(2), 129–157. doi:10.1002/cne.901750202.
- Wittgenstein, L. (1953). *Philosophical investigations*. Translated from German by G.E.M. Anscombe (3rd edition, 1986). Oxford: Basil Blackwell.
- Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, 24(1), 11–17, 111–125.

- Womelsdorf, T., Schöffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., et al. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science*, 316(5831), 1609–1612. doi:10.1126/science.1139597.
- Woodward, T. S., Menon, M., & Whitman, J. C. (2007). Source monitoring biases and auditory hallucinations. *Cognitive Neuropsychiatry*, 12(6), 477–494. doi:10.1080/13546800701307198.
- Woody, E. Z., & Szechtman, H. (2011). Adaptation to potential threat: the evolution, neurobiology, and psychopathology of the security motivation system. *Neuroscience and Biobehavioral Reviews*, 35(4), 1019–1033.
- Woolf, N. J., & Hameroff, S. R. (2001). A quantum approach to visual consciousness. *Trends in Cognitive Sciences*, 5(11), 472–478.
- Worsley, C. L., Recce, M., Spiers, H. J., Marley, J., Polkey, C. E., & Morris, R. G. (2001). Path integration following temporal lobectomy in humans. *Neuropsychologia*, 39(5), 452–464.
- Wu, S., & Amari, S. (2005). Computing with continuous attractors: stability and online aspects. *Neural Computation*, 17(10), 2215–2239.
- Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research*, 48(20), 2070–2089. doi:10.1016/j.visres.2008.03.021.
- Wyatte, D., Curran, T., & O'Reilly, R. (2012). The limits of feedforward vision: recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, 24(11), 2248–2261. doi:10.1162/jocn_a_00282.
- Wyss, R., König, P., & Verschure, P. F. (2006). A model of the ventral visual system based on temporal stability and local memory. *PLoS Biology*, 4(5), e120. doi:10.1371/journal.pbio.0040120.
- Xing, J., & Andersen, R. A. (2000). Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames. *Journal of Cognitive Neuroscience*, 12(4), 601–614.
- Xu, F., Greer, C. A., & Shepherd, G. M. (2000). Odor maps in the olfactory bulb. *Journal of Comparative Neurology*, 422(4), 489–495.
- Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., & Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*, 11(11), 1352–1360.
- Yeterian, E. H., & Van Hoesen, G. W. (1978). Cortico-striate projections in the rhesus monkey: the organization of certain cortico-caudate connections. *Brain Research*, 139(1), 43–63.
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, 111(4), 931–959.
- Yim, M. Y., Aertsen, A., & Kumar, A. (2011). Significance of input correlations in striatal function. *PLoS Computational Biology*, 7(11), e1002254. doi:10.1371/journal.pcbi.1002254.
- Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M., & Deisseroth, K. (2011). Optogenetics in neural systems. *Neuron*, 71(1), 9–34. doi:10.1016/j.neuron.2011.06.004.
- Yoshinaga, H., Ohtsuka, Y., Tamai, K., Tamura, I., Ito, M., Ohmori, I., et al. (2004). EEG in childhood absence epilepsy. *Seizure*, 13(5), 296–302. doi:10.1016/S1059-1311(03)00196-1.
- Young, M. P., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, 256(5061), 1327–1331.
- Young, T. (1802). Bakerian Lecture: On the theory of light and colours. *Philosophical Transactions of the Royal Society of London*, 92, 12–48.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), 547–558.
- Zeki, S. (1993). *A vision of the brain*. London: Blackwell.
- Zeki, S. (2001). Localization and globalization in conscious vision. *Annual Review of Neuroscience*, 24, 57–86.
- Zeki, S. (2005). The Ferrier Lecture 1995: Behind the seen: the functional specialization of the brain in space and time. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1458), 1145–1183. doi:10.1098/rstb.2005.1666.

- Zeki, S. (2008). The disunity of consciousness. *Progress in Brain Research*, 168, 11–18. doi:10.1016/S0079-6123(07)68002-9.
- Zeki, S., & Bartels, A. (1999). Toward a theory of visual consciousness. *Consciousness and Cognition*, 8, 225–259.
- Zelano, C., & Sobel, N. (2005). Humans as an animal model for systems-level organization of olfaction. *Neuron*, 48(3), 431–454. doi:10.1016/j.neuron.2005.10.009.
- Zeman, A. (2001). Consciousness. *Brain*, 124(Pt 7), 1263–1289.
- Zhang, Z. W. (2004). Maturation of layer V pyramidal neurons in the rat prefrontal cortex: intrinsic properties and synaptic function. *Journal of Neurophysiology*, 91(3), 1171–1182. doi:10.1152/jn.00855.2003.
- Zhong, W., Gallivan, J. P., Zhang, Y., Li, L., Lester, H. A., & Dougherty, D. A. (1998). From ab initio quantum mechanics to molecular neurobiology: a cation-pi binding site in the nicotinic receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21), 12088–12093.
- Zhou, Y. D., & Fuster, J. M. (2000). Visuo-tactile cross-modal associations in cortical somatosensory cells. *Proceedings of the National Academy of Sciences of the United States of America*, 97(17), 9777–9782.
- Zihl, J., von Cramon, D., & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain*, 106(Pt 2), 313–340.
- Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331(6158), 679–684. doi:10.1038/331679a0.
- Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370(6485), 140–143. doi:10.1038/370140a0.

Index

- Aboutness. *See* Intentionality
Abstraction, 294, 302–304
Access. *See* Consciousness
Acetylcholine, 14, 28, 178, 181, 189–191
Achromatopsia, 71, 73, 76, 81, 295
Action potential, 11, 16, 20, 22, 83, 103, 108, 137, 164
Adaptive resonance theory. *See* Grossberg, S.
Address, 41, 90, 97, 218, 311. *See also* Content addressable
Ageusia, 78
Aggregate, 25, 108, 116, 266, 267, 269, 271, 273
Agnosia. *See also* Prosopagnosia; Simultanagnosia
auditory, 79
somatosensory, 79
visual, 73–76, 211, 225, 227
Agranular cortex, 136, 147
AI. *See* Artificial Intelligence
Akinesia, 159, 160
Akinetic mutism, 159
Akinetopsia, 69, 76, 225–227, 252
Algorithm, 86, 89, 114, 266, 273, 291
Alignment, 144, 213, 231, 234, 235, 306
Allocentric, 33, 212, 213, 235, 238. *See also* World-referenced
Ambiguity, perceptual, 94, 96, 97, 100, 218, 222, 223, 233, 253, 261
Amnesia, 58, 59, 66, 76
AMPA receptor, 16, 23, 189
Amphibian, 309
Amphioxus. *See* Lancelet
Amygdala, 25, 35, 44, 59–61, 147, 157, 182, 254, 255
Anesthesia, 19, 114, 180, 181, 189–194, 197, 201, 204, 210, 313
Anesthetic
general, 189, 190, 193, 194, 197, 210, 229, 319
local, 16, 57, 128
Aniston, Jennifer, 253, 254
Anosmia, 78, 149
Ant, 283, 304, 305
Anterior temporal lobe, 259
Anticorrelated. *See* Correlation
Anton-Babinski syndrome (visual anosognosia), 300
Aphasia, 66
Appraisal, 54, 60
Arbitrary. *See* Representation
Arousal
nuclei controlling, 28, 29, 164, 178, 192, 313
system, 178, 258
Artificial Intelligence, 118, 120, 297, 312, 318
Ascending projections (cortical), 138–140, 144, 147, 169, 171. *See also* Hierarchy
Assembly, 83, 100, 183, 188, 189, 204, 207, 245, 253, 269, 277
Association, 9, 52, 116, 132, 183, 251
Associative
dynamic associative (*see* Dynamic associative network)
learning, 23, 225
memory, 57, 90, 96–98, 147, 268
meaning, 9, 225, 259
recall, 53, 58
Ataxia, 28, 150
Atomic consciousness, 127, 128
Attention
attentional focus, 210, 223, 241, 242, 258
attentiveness, 160, 176, 203, 204
and binding problem, 241, 243
spatial, 210, 241–243
Attractor
basin of attraction, 98, 261, 262
continuous, 109, 261, 306
dynamic attractor network (*see* Dynamic associative network)
multimodal, 249, 261, 268
stability, 99, 261
switching between states, 262
unimodal, 249, 261
Auditory cortex
primary auditory cortex (A1), 31, 32, 36, 38, 39, 42, 78, 141, 145, 146, 190, 229, 314
secondary auditory cortex (A2), 32, 78, 145, 229

- Autoassociative network, 94, 96
 Automaticity, automatic response, 7, 135, 156, 293, 321
 Automatism. *See* Epilepsy, absence
 Autonomous nervous system, 27, 35, 53, 60–63, 167
 Autoscopic experience, 213, 320
 Autostereogram, 216
 Awareness, 4, 76, 103, 132, 167, 188, 195, 198, 285, 296, 317
 Axon, 12, 13, 19, 20, 40, 84, 85, 88, 96, 136, 138, 289
- Baars, B.J., 118, 120
 Back-propagation algorithm, 89
 Bálint's syndrome, 242
 Basal ganglia, 30, 31, 156–168, 181, 289, 293, 309
 Bat, 7, 78, 146, 309, 310
 Bayesian model, 260
 Behavioral inhibition, 160, 183, 203. *See also* Disinhibition
 Behaviorism, 55, 84, 292
 Belief, 56, 106, 222, 286, 289, 291
 Bell, C., 83, 103
 Berkeley, G., 297, 303
 Beta. *See* Oscillation
 Binding. *See also* Attention, and binding problem
 of features, 64, 125, 201, 211, 231, 233, 239–244, 248, 284
 by location, 242
 problem, 50, 201, 239–241
 Binocular rivalry, 188, 198, 199, 216, 262
 Biological clock. *See* Suprachiasmatic nucleus
 Bistability, 94
 Blink
 attentional, 199
 eye, 51, 198, 199, 262
 Blindness, 70, 76, 149, 150, 200, 217, 300
 blind-born, 215
 blindsight, 73, 214
 color-blind, 71 (*see also* Achromatopsia)
 Block, N., 116, 118
 Blood-oxygenation level dependent (BOLD) signal, 195, 199, 319
 Body
 body-centered, 40, 236
 language, 4, 66, 73, 112
 map (*see* Map, body)
 orientation, 235
 ownership, 229
 position, 31, 49, 133, 213
 posture, 28, 31, 133, 151, 159, 213
 Brain stem, 27–30, 153, 155, 156, 164, 168, 178, 190
 Broca's area, 66, 149
- Brodmann areas (general), 34, 317
 area 1 (*see* Somatosensory cortex)
 area 2 (*see* Somatosensory cortex)
 area 3a, 3b. (*see* Somatosensory cortex)
 area 4 (*see* Motor cortex, primary)
 area 5 (*see* Parietal cortex)
 area 7A or 7 (*see* Parietal cortex, posterior)
- CA1, CA3. *See* Hippocampus
 Calcium, Ca^{2+} , 18, 22, 180, 190, 191
 Calibration, 151, 182, 239
 Great Calibrator, 237
 Camera, 36, 40, 43, 117, 133, 213, 283
 Capgras, syndrome of, 62, 63
 Caravaggio, 133–135
 Cardinal cells, 239, 241, 245
 Cartesian. *See* Descartes
 Category
 classify into, 87, 111, 117, 133
 coding, 253, 254, 258
 decision, 240
 mistake, 199, 268
 perceptual category, 74, 120, 223, 255, 256, 302
 Caudate nucleus. *See* Striatum
 Causal interaction, 281, 284, 288, 290, 296
 Causal intermediate, 55, 56, 62, 81, 291, 292, 317
 Causation, upward or downward, 280
 Cell
 body, 11–13, 20–22, 84, 85, 96, 136, 174, 191
 conceptual or concept cell, 253–257
 Central processing unit (CPU), 90, 91
 Cerebellum, 3, 25, 26, 28, 150–156, 168, 201, 234, 293
 C-fiber, 289, 290, 321
 Chalmers, D., 49, 81, 127, 283–288, 295, 296
 Channelrhodopsin, 245
 Charcot–Wilbrand's syndrome, 180
 Chess, 87, 310
 Chinese
 nation, 116
 room, 107, 108, 291, 297
 Chip, electronic, 287, 288, 292, 297, 313, 314
 Cholinergic. *See* Acetylcholine
 Churchland, Patricia, 6, 100, 104
 Churchland, Paul, 6, 100, 222
 Cingulate cortex, 34, 35, 54, 60, 118, 149, 150, 157, 159, 160, 167, 182, 193, 223
 Classification, 88–91, 94, 133, 218, 253, 258
 Claustrum, 150
 Coactivation, 90, 111, 248, 276
 Coalition, 188, 207
 Code
 firing rate code, 184, 244–249, 252, 262, 263, 270
 phase code, 184, 244, 249, 251, 252, 262, 270, 311
 (*see also* Spike, phase)
 relational, 252, 262, 270, 271
 sparse, 254, 255

- Collinearity, 94, 95, 98, 242, 260
Color
color perception or color vision, 71–73, 102, 219, 262, 278, 279, 281, 295
color space, 279, 302
constancy, 223, 224, 320
discrimination, 73, 102
information processing, 71, 143
Column (cortical), 136, 138, 142–145, 156, 168, 170, 172, 201
Coma, 29, 128, 164, 213, 312, 313
Combinatorial
explosion, 241, 245
scheme, 239, 240
Common fate, 94, 95
Complexity
of information, 120, 124–127, 163, 313
representational, 172, 281, 306, 312, 321
Computational functionalism, 114–117
Computer program, 92, 107, 297
Conceivability, 286–288, 314
Concept cell. *See Cell*
Cone, cone receptor, 71, 102, 302
Congruency, multimodal, 248
Connection strength, connection weight, 88, 89, 97, 99, 110–112, 237
Connectivity
functional, 166, 193, 270, 274, 319
multimodal, 171, 248, 249
Consciousness
access, 6, 118, 119, 318
in a bottle, 252
definition of, 2, 62
easy problems, 49, 81, 295, 296
enabling factor, 178, 190, 191
as epiphenomenon, 135, 281, 282, 286–289
hard problem, 49, 81, 295, 296
loss of, 70, 149, 150, 159, 183, 189–193, 234, 309
marker of, 309, 313
and motor behavior, 62–69, 81, 164, 292
as multimodal phenomenon, 49, 167, 171, 217, 220, 271, 272, 281
neural correlate of, 149, 259, 260, 283, 284, 290
neurocomputational account of, 100
phenomenal, 6, 118, 119, 209, 318
as potentiality, 126, 293
richness, 59, 106, 125, 126, 210, 220, 263, 271, 290, 305, 307
stream of, 109, 187, 205, 260
unity of, unified nature of, 121, 132, 231, 297
Consistency
crossmodal, 302, 303
similarity-based, 278
temporal, 268, 278, 279, 284, 303
Construction, constructive, 43, 45, 56, 125, 132, 160, 168, 171, 186, 216, 222, 223, 235, 237, 244, 247, 256, 268, 288, 299, 302, 314
Content addressable, 97, 98
Context
action and task, 92, 159, 210
bodily, 167
color, 223–225, 259, 271
multimodal, 273
spatial and temporal, 33, 45, 53, 157, 167, 212, 242, 299
Continuity, 51, 52, 59, 94, 109, 201, 205, 212
Control
cognitive, 149, 150, 229, 314
layers, 30
of movement, 28, 31
top-down, 141, 258, 262, 307
Convergence
multimodal, 171, 234, 271, 306
in neural network model, 97, 98
Corpus callosum, 26, 31, 35, 67
Correlation, 90, 128, 184, 186–191, 199, 201, 219, 246–248
anticorrelation, 96, 97, 246, 247
decorrelation, 184, 190, 207, 247
noise correlation, 184
Correspondence, 272, 281–284, 289, 292–295
Corticocortical, 118, 136, 139, 143–145, 156, 168–170, 192, 197
Corticothalamic, 100, 120, 135, 136, 141, 156, 166–168, 174, 273. *See also Thalamocortical*
Cranial nerve, 27, 105, 107
Craniotopic. *See Head; Map*
Crick, F., 149, 188, 230
Critic, 89
Cross-modal, 171, 220, 221, 237, 247, 259, 302, 303.
See also Multimodal
Cuneiform room, 107, 108, 297, 299
Cyclone, 126, 220, 263
Decoding, 111, 218, 266, 274, 277. *See also Read out*
Decorrelation. *See Correlation*
Deep-brain stimulation, 47, 150, 191, 192, 313
Default mode network, 193
Degeneration, 159, 250, 314
Degradation, 96, 98
Dehaene, S., and Changeux, J.-P., 118–120, 169, 195, 197
Delusion, 62, 149, 229
Dendrite, 11–14, 20–23, 84, 96, 136, 138, 174, 180
Dennett, D.C., 49, 51, 56, 81, 114, 230
Depolarization, 15–18, 23, 28, 84, 85, 174, 176, 178, 200
Depth
illusory, 262
perception, 102, 210, 216
Damasio, A., 159, 228, 230
Day-night clock. *See Suprachiasmatic nucleus*
Department store, 269

- Descartes, R., 7, 38, 56, 84, 91, 120, 173, 231, 280, 286, 302
- Descending projections (cortical). *See* Ascending projections
- Desynchronized EEG, 175, 176, 178, 182–184, 207, 260. *See also* Electroencephalography
- Development
- evolutionary, 27, 66
 - ontogenetic, 41, 42, 228, 237–239, 247, 248, 278, 320
- Diencephalon, 25, 28, 29
- Differentiation, 121, 124, 126, 166, 220, 222
- qualitatively differentiated, 49, 81, 106
- Digital computer, 83, 107, 114, 222
- Dimension. *See* Sensory dimension
- Direct realism, 297–300
- Discriminative state, 56, 62, 71, 73, 81, 106, 252, 291
- Disinhibition, 160, 161, 163, 182. *See also* Behavioral inhibition
- Disparity. *See* Stereodisparity
- Disposition, 56, 106, 126, 247, 291
- Dopamine
- dopaminergic cell, 18, 90, 158, 159, 162, 163
 - neuromodulation and state setting, 14, 28
- Dormant, 113, 293
- Down state. *See* Up and down states
- Dream, 113, 173, 178–184, 188–190, 200, 293, 300
- Dualism, 38, 56, 230, 280, 281, 285, 286, 288, 289, 295, 296, 300
- Duck and rabbit. *See* Jastrow
- Dynamic associative network, 249, 261, 263, 268. *See also* Attractor
- Echoing, 256, 259, 271, 273, 306, 311
- Edelman, G. M., 120–125, 162, 163, 313
- Efference copy, 155, 219, 235
- Egocentric
- perspective (*see* Perspective, first-person reference frame (*see* Frame of reference))
- Electroencephalography (EEG), 21, 68, 125, 150, 173–185, 189–191, 194–196, 204, 205, 313, 319
- alpha, beta, delta, gamma, theta (*see* Oscillation)
- Electromagnetic, 43, 127, 133, 292
- Emergence, emergent property, 92, 97–99, 120, 127, 181, 211, 260, 268, 280, 321
- Emotion, 53, 56, 59, 179
- emotional behavior, 35, 56, 60, 67, 149, 159
 - emotional feeling, 54, 60, 62, 160, 227 (*see also* Feeling)
 - emotional value, 35, 54, 60, 148, 157, 162, 184, 213, 277
 - memory (*see* Memory)
- Enactment, 276, 277
- Encephalitis, 54, 159, 254
- Energy (in a neural network), 97, 98, 262
- Engram, 91
- Ensemble, 184, 185, 190, 191, 221, 247, 260, 269–278, 281, 284, 293, 294, 306, 319
- Entorhinal cortex, 30, 33, 34, 57, 179, 254
- Entropy, 120–124, 126, 296
- Epilepsy
- absence (petit mal), 67, 68, 179, 183
 - patient, 57, 150, 176, 204, 213, 253, 255, 299
 - seizure, 67, 68, 78, 125, 204
 - single-cell recordings, 253–255
- Epiphenomenon. *See* Consciousness
- Epistemology, 283, 295, 303
- Equipotential, 41
- Error, 35, 89, 91, 151, 153–155, 162, 220
- Essential node. *See* Node
- Event-related potential (ERP), 195–197, 199, 206
- Evolution, 27, 63–66, 131, 132, 148, 166, 217, 307
- Excitatory postsynaptic potential (EPSP), 17, 21–24, 174, 203
- Executive
- function, 118–120, 134
 - summary, 133
- Expectation, 35, 52, 151, 203, 204. *See also* Prediction
- Experiential content, 4, 49, 128, 178. *See also* Phenomenal content
- Explanatory gap, 9, 265, 266, 277, 283
- Externalism, 292, 293, 301, 315, 322
- Extrastriate cortex, 29, 36, 69, 70, 74, 136, 141, 197, 198, 258
- Eye
- blink, 262
 - eye-centered, 234–237
 - eye-hand coordination, 237
 - movement, 28, 30, 39, 50, 66, 135, 138, 153, 155, 161, 173, 237, 309
 - position, 234, 236, 237
 - saccade, 30, 50, 51, 211, 262
- Face, facial
- expression, 66, 76, 159, 210, 311
 - features, 42, 90, 210
 - recognition, 63, 74, 76, 90–96, 253, 254
- Familiarity, 52, 58, 62, 73–76, 227, 260
- Fault tolerance, 98, 245
- Fear, 25, 44, 54, 59, 60, 160
- Feature conjunction, 243
- Feature detector, 146, 171, 221, 239, 241, 244–253, 258, 261
- Feedforward wave. *See* Forward sweep
- Feeling, feel, 48, 54–56, 60–62, 79, 102, 106, 129, 133, 160, 193, 214, 222, 227, 231, 287, 291, 293
- Figure-ground (figure against background), 94, 95, 199, 201
- Filter, 141, 153–156, 307
- Firing. *See* Spike
- Fodor, J., 115, 118, 283
- Folk psychology, 6, 100, 289

- Forward sweep, 139, 195, 197
Fovea, 36, 37, 39
Frame
 coordinate, 234
 of reference, 39, 40, 212, 213, 234, 235, 237, 320
Free will, 63
Frontal eye fields, 30, 66, 138, 157
Frontal lobotomy, 149
Frontoparietal network, 197, 198, 206
Functionalism, 55, 56, 62, 71, 73, 81, 106, 108, 110, 114–117, 135, 291, 292, 297, 314
Functional magnetic resonance imaging (fMRI), 185, 195, 199, 229, 312, 319
Fungus, 304, 305
Fusiform
 face area, 199
 gyrus, 71, 74, 76, 197, 248
- Gamma-aminobutyric acid (GABA), 14, 16–22, 57, 104, 138, 157, 163–166, 184, 189, 280, 318
Gaze, 226, 234, 235
Gestalt
 feature, 95, 201, 242
 grouping. *See* Grouping
 property. *See* Gestalt, feature
 psychology, 94, 98, 211, 296
 as a whole, 98, 270
Gist, 132, 133, 149, 312
Glacier, 126–129, 296
Global workspace, 118–120, 169, 188, 197
Globus pallidus (pallidum), 31, 157–160, 168
Glutamate, 14, 16, 17, 22, 23, 104, 136, 174, 180, 181, 189
Goal, 54, 64, 89, 150, 186, 266. *See also*
 Representation
 goal-directed behavior, 61, 64
Graceful degradation. *See* Degradation
Grandmother cell, 239–241, 254
Grasp (of object), 69, 144, 259, 310
Gray, C. M., 200–202
Grossberg, S., 258
Grouping
 attention-based, 244
 contour, 244
 feature, 94, 95, 211, 231, 242
 memory-guided, 242
 pre-attentive Gestalt, 242, 244
Gustatory, 27, 41, 78, 148. *See also* Insular cortex
Gut, 53, 60, 149, 167, 214. *See also* Visceral
- Hallucination, 78, 160, 178, 182, 229, 230, 299, 314, 320
Head
 direction, 33
 head-centered, 40, 156, 234–237
 orientation, 146, 211, 238
 rotation, 28, 40, 153–155, 213, 238
- Hear, hearing, 38, 78, 149, 229, 300, 310
Heartbeat, 35, 53, 60, 164
Hebb, D., 23, 83, 90, 95, 111, 268, 269
Helmholtz, H. von, 103, 215
Hemianesthesia, 79, 214, 215
Hemineglect, 76, 77, 79, 150, 159
Hierarchy, 32, 33, 79, 138–147, 149, 168, 169, 197–200, 227, 240, 253, 269, 306
Hippocampal formation, 39, 58, 61, 255. *See also*
 Memory
Hippocampus, 14, 22–25, 33, 34, 57, 139, 176, 178–188, 254, 259, 270, 317
H.M. *See* Molaison, H.
Homeostasis, 29, 228, 230
Homology, 306, 307, 309
Homunculus
 in brain, 38, 120, 149, 230, 277
 map, 39, 45, 235
Hopfield, J. J., 87, 95, 97, 99, 112, 268
Horizontal connections, 136, 139, 169, 170
Hume, D., 228, 297
Huntington's disease, 159
Hyperpolarization, 15–19, 174
Hypothalamus, 18, 28–30, 61, 163–168, 178
- Idealism, 303
Identification of sensory modality. *See* Modality identification
Identity
 personal, 228, 254, 256
 theory, 289, 290
Illusion, illusory, 95, 102, 224–226, 231, 235, 237, 260–262, 276, 298–302
Imagery, 76, 81, 113, 135, 178, 179, 182, 187, 217, 230, 260, 293, 313
Imagine, imagination, 4, 7, 44, 113, 167, 212, 225, 255–260, 269, 276, 282–288, 304
Immersion, immersive, 133, 210, 231, 306
Impenetrability, 167, 282, 283, 294
Impression (sensory impression, sense impression), 83, 101, 215, 226, 298
Imprinting, 95, 96, 112
Inanimate systems, 10, 107, 109, 112, 127, 220, 244, 263
Individuation, 247, 252, 258, 292. *See also* Modality identification
Inferotemporal (IT) cortex, 32, 50, 70, 71, 74, 75, 139, 188, 240, 241, 248, 258, 260
Inferior olive, inferior olivary nucleus, 151–154
Inferior temporal. *See* Inferotemporal cortex
Infinite regression, 116, 268
Information theory, 120–127, 263, 296
 amount of information, 121, 133, 163
 complexity of information (*see* Complexity)
 integration of information (*see* Integration)

- Inhibition
 behavioral (*see* Disinhibition)
 cellular (*see* Inhibitory postsynaptic potential)
- Inhibitory postsynaptic potential (IPSP),
 17, 18
- Insect, 302–304, 307
- Insular cortex, Insula, 34, 35, 54, 78, 79, 171
- Integration
 feature, 239, 244
 of information, 120, 123, 139, 163, 168, 211
 multimodal, 50, 51, 210, 306–310
 multisensory, 31, 237, 307, 308, 311
 sensory, 49, 105, 108, 244, 307
- Intelligence, 129, 310, 311, 318. *See also* Artificial Intelligence
- Intentionality, 8, 9, 116, 117, 220, 222, 298, 301
- Internet, 120, 285, 311
- Interneuron, 21, 22, 96, 138, 151, 163, 200, 280
- Interoception, 48
- Interpretation
 cognitive, 94, 223–227, 233, 253–259, 262, 273, 306, 312
 interpreter system, 246, 247, 262
 linguistic, 67, 256
 perceptual, 94, 118, 225, 260, 271, 273, 277
 of sensory inputs, 38, 104, 195, 217, 218, 223–225, 231, 259, 276–278
- Introspection, 5, 55, 228
- Invariance, Invariant, 90, 117, 153, 240
- Inverted spectrum argument (inverted color perception), 278, 279, 291
- Irreducibility, 287, 292, 304, 315
- Isomorphism, 274, 275, 296
- Itch, 50, 51, 237
- Jackson, F., 294, 295, 314
- James, W., 51, 109
- Jastrow's duck and rabbit, 94, 133, 216, 223, 225, 253, 261, 320
- Jellyfish, 307
- Jeopardy, 310, 311
- Kandel, E. R., 22, 43, 168
- Kanizsa triangle, 95, 260, 299
- Kant, I., 132, 167, 212, 239, 241, 298
- Ketamine, K-hole, 189, 193, 229
- Kinesesthesia, 50, 51, 299, 317
- Koch, C., 127, 132, 149, 188, 230, 312
- Labeled-lines hypothesis, 101–105, 233, 278, 290
- Lancelet, 26, 27
- Language, 4, 8, 49, 62–69, 73, 81, 100, 114, 115, 220–222, 256, 268, 282, 290, 299, 301, 302
- Lateral geniculate nucleus (LGN), 18, 28, 36, 40, 70, 73, 91, 137, 139, 141, 169, 178, 226
- Learning
 associative (*see* Associative learning)
 neuromodulation of, 28
 rule, 86, 89, 90, 92, 95, 107
 spatial, 25
 supervised, 89
 unsupervised, 89, 90, 111
- Leech, 307
- Level
 aggregate, 108, 266, 267, 269, 271, 273, 288 (*see also* Aggregate)
 representational, level of representation (*see* Representation)
- Light, wavelength of, 43, 71, 102, 103, 268, 277, 294
- Linguistic. *See* Language
- Llinás, R., 63, 64
- Local field potential (LFP), 21, 174, 176, 179, 183, 201–203, 308
- Locke, J., 7, 49, 101, 102, 278, 297
- Locked-in, 67, 68, 112, 293, 313
- Locust, 307, 308
- Long-term depression (LTD), 22, 86, 91, 151, 153, 155
- Long-term potentiation (LTP), 22, 24, 25, 86, 91, 151, 155, 162
- Loop (neocortex, basal ganglia, thalamus), 157–159, 162, 163, 168
- Mao Zedong, 92, 93, 95–98, 100
- Magnetoencephalography (MEG), 200, 204, 205, 229
- Magritte, R., 104, 279
- Map, mapping
 body, 36, 39, 45, 214, 215, 230, 231, 237
 craniotopic, 40, 156
 egocentric, 235
 input-output, 89, 100
 multimodal, 39, 234
 retinotopic, 36, 39, 42, 156, 237
 sensory, 35, 38–40, 156, 230, 249, 262, 270
 spatial, 42, 233–237, 242, 248
 tonotopic, 36, 38, 78, 309
 unimodal, 249, 261
- Marr, D., 10, 132, 142, 151, 153, 230, 266, 284
- Mary the Color Scientist, 294, 295
- Materialism
 eliminative, 100, 108, 289, 290, 315
 in general, 107, 286–289, 294, 297, 298, 315
- McCulloch, W., and Pitts, W., 83–85
- Meaning
 associative (*see* Associative meaning)
 concept of, 8, 9, 225
 conceptual meaning, 133, 145
 generation of, 245, 278
 phenomenal (*see* Phenomenal meaning)

- Mechanoreceptor, 79, 88, 144, 213, 302
Medial geniculate nucleus, 42, 141, 145
Medial temporal lobe (MTL), 57–59, 62, 223, 227, 253, 254, 258–262
and categorization, 32, 223, 253, 258, 259
Medulla, 25–27, 30, 79, 164, 178
Membrane potential. *See* Voltage, membrane
Memory
autobiographical, 9, 25, 58 (*see also* episodic)
consolidation, 182, 184, 187, 189
declarative, 58, 59, 217, 227, 314
emotional, 44, 59, 217
encoding, 91, 98, 149, 203
episodic, 58, 182, 184, 209
formation, 33, 57, 159, 182, 184, 187
hippocampal, 33, 58, 184
iconic, 211, 319
long-term, 25, 33, 52, 53, 57–59, 178, 212
procedural, 53, 58
retrieval or recall, 9, 58, 63, 68, 90, 91, 97, 113, 179–182, 193, 242, 259, 314
semantic, 58, 223, 256–259
short-term, 52, 100, 178, 212
trace, 76, 242, 314
working (*see* Working memory)
Mental
phenomena, 100, 289, 290, 315
properties, 55, 286, 300
state, 81, 107, 222, 291
Merleau-Ponty, M., 47
Mesencephalon, 18, 25, 30, 158, 163
Meta-network, 271–273, 276, 277, 281–284, 293, 294, 306, 321
Microconsciousness, 127, 128
Microgenesis, 195, 207, 215
Midbrain. *See* Mesencephalon
Middle temporal area (MT). *See* V5
Mirrors (metaphor of echoing mirrors), 254–258, 263.
See also Echoing
Misrepresentation, 97, 299, 301
Mnemonic, 150, 225, 242, 253, 255, 263. *See also*
Memory
Modality
identification, 101–107, 117, 218, 220, 233, 241, 247, 248, 258, 261, 268, 306, 310
self-identification, 220, 221, 247, 263
sensory, 29, 33, 39–42, 48–50, 100–106, 133, 142, 171, 193, 217, 234, 247, 272, 292
submodality, 48, 49, 69–71, 78, 79, 101, 102, 142, 156, 210, 231, 247–249, 252, 271, 278, 281
topology of sensory modalities, 219, 246, 247
Model. *See* World
Molaison, H., 57, 58, 76
Monism, 281, 286, 315
Motion vision, 50, 51, 69–71, 81, 128, 201, 226, 227, 252, 271
Motivation, 28, 148, 157, 159
Motor cortex (primary, secondary, supplementary), 31, 66, 80, 157, 158, 204, 270
premotor (*see* Premotor cortex)
Müller, J., 103
Multilevel, 266, 269, 273, 283, 284, 290, 295, 304, 312, 315. *See also* Representation, representational levels
Multimodal
brain area, 171, 193, 234
cross-talk, 171, 250
feedback, 258
integration, 9, 50, 51, 132, 210, 307, 310
mixing, 250
stack, 248, 251, 259, 261
Multiple personality disorder, 229, 313
Multiple realizability, 56, 116, 292, 296, 297
Multisensory. *See also* Integration
domain, 272, 273
sensation, 52
systems, 263
Muscimol, 57–59, 66, 69
Mutual information, 123–125, 163
 Na^+ channel, 15, 16, 128, 137
Nagel, T., 7, 309
Naïve realism. *See* Direct realism
Naming game, 246, 247
Nasal epithelium, 30, 78, 101, 148, 214
Navigation (spatial), 33
Neocortical layers (laminae), 19–22, 136–140, 142, 144–148, 170, 171, 181, 192, 203, 317
Neural correlate of consciousness. *See* Consciousness
Neural network, neural net
activity, 127, 288
model, 54, 83, 86, 90–98, 100, 104, 111–113, 116, 156, 168, 209, 237, 263, 266, 268, 273, 288
Neurorepresentationalism. *See* Representationalism
Newton's color circle, 278, 279
NMDA receptor, 23, 25, 189
Nociception. *See* Pain
Node
essential node, 71, 79, 81, 149, 150, 211, 231, 252, 259, 276
network node, 115, 116, 120, 169, 170, 268
Noise, noisy, 93, 97, 98, 270
noise correlation (*see* Correlation)
Norepinephrine (noradrenalin), 28, 176, 178
Nucleus accumbens, 60. *See also* Striatum
Numerosity, 101, 102
Occipitotemporal areas, 74, 171, 198, 199
Odor. *See* Smell
Odorant. *See* Smell
Odor receptor. *See* Smell
Olfaction. *See* Smell

- Olfactory
 bulb, 30, 78, 146–148
 receptor, 101, 308
- Ontogeny, 41, 81, 228, 230, 278, 284
- Optogenetics, 47, 138, 141, 144, 203, 313
- Orbitofrontal cortex. *See* Prefrontal cortex
- Orgasm, 35, 293
- Orientation
 of body (*see* Body)
 of head (*see* Head)
 orientation behavior, 39, 171, 176, 309
 tuning, 170
 of visual stimuli, 31, 42, 71, 90, 136, 170, 201, 240, 271
- Oscillation
 alpha, 21, 200, 204–207
 beta, 21, 200, 203–207, 251
 delta, 21, 174–176, 192, 200
 gamma, 21, 176, 183, 200–204, 207, 249, 260, 280, 283, 284
 in general, 19, 164, 165, 191, 203, 251
 theta, 21, 176, 178, 182, 185, 205
- Osmolality, osmosensor, 29, 163, 167
- Out-of-body experience, 213, 229, 320
- Pain perception or sensation, 35, 45, 48, 54, 129, 142, 143, 150, 160, 289, 290, 321
- Pallidum. *See* Globus pallidus
- Panspsychism, 5, 10, 109–116, 118, 120, 125–129, 166, 209, 263, 280, 281, 296, 315, 318
- Parahippocampal
 cortex, 145, 182, 258
 gyrus, 34
 place area, 199
- Parallel-distributed, 91, 98, 241
 processing (PDP), 90, 91, 120
- Paralysis, 66, 67, 292, 293
- Paramecium, 64, 65
- Parietal cortex
 Brodmann's area 5, 31, 142, 144, 259
Brodmann's area 7A or 7 (see posterior parietal cortex)
 and categorization, 258
 inferior parietal, 77, 171, 193, 234
 lateral intraparietal (LIP), 171, 234
 parietal lobe, 31, 32, 59, 76
 posterior parietal cortex, 30–32, 38–40, 76, 141, 142, 144, 150, 223, 234, 259, 317
 ventral intraparietal (VIP), 171, 234
- Parkinson's disease, 18, 47, 158, 159, 191, 192
- Pattern completion, 93, 98, 112
- Penfield, W., 285, 299
- Perceptron, 86–88
- Perceptual
 decision, 32, 98, 128, 244
 interpretation (*see* Interpretation, perceptual)
- Perirhinal cortex, 34, 58, 227, 258
- Perspective
 first-person, 33, 39, 40, 49, 121, 132, 211–213, 234, 235, 238, 297
 perspectival unity, 50, 212, 213, 229, 230, 234
 visual, 8, 320
- Phantom
 pain, 45
 sensation, 293
- Phase
 code (*see* Code, phase)
 locking, 201, 205, 206
 shifting, 251, 252, 321
 spike (*see* Spike, phase)
- Phenomenal
 content, 6, 118, 125, 178, 211, 220–222, 225, 274, 291
 experience, 4, 6, 8, 10, 99, 220, 266–268, 273, 283–287, 290–292, 298, 314
 meaning, 9, 40, 225, 233, 241, 244, 248, 259, 262, 268, 271, 284
- Phenomenology, phenomenological experience, 49, 230, 266, 268, 294, 295, 303
- Phosphone, 293, 300
- Photoreceptor, 36, 43, 64, 71, 91, 103, 282, 294
- Physicalism. *See* Materialism
- Physicochemical property, 107, 276, 284
- Piriform cortex, 30, 78, 146–148, 203
- Pitch, 31, 38, 39, 78, 146
- Planning, 6, 41, 49, 64–66, 120, 135, 149, 157, 186, 203, 305
- Plasticity. *See* Synaptic plasticity
- Pleasure, pleasant, 35, 54, 160
- Polysensory. *See* Multimodal
- Pons, pontine nuclei, 26–30, 152, 156, 178
- Population
 activity, 174, 186, 199
 code, 237, 254
- Potential, membrane. *See* Voltage, membrane
- Potentiality. *See* Consciousness
- Preconscious, 118, 223, 228
- Prediction
 predictive coding, 204, 220, 261
 reward, 160, 162, 203
 sensory, 151, 156, 171, 204, 220, 261
- Prefrontal cortex
 anatomy of, 30, 31, 41, 60, 61, 138, 144, 147, 158, 166
 and categorization, 223, 253, 258
 dorsolateral, 149, 157, 234
 general, 118, 141, 149, 150, 182, 184–186, 193, 229, 314
 medial, 60, 147, 157, 159, 185, 187, 229
 orbitofrontal, 34, 60, 78, 146–150, 160
 and working memory, 53, 59, 317
- Premotor cortex, 66, 76, 77, 157, 158, 171, 197, 204, 234, 312

- Preoptic area, 29, 164, 165
Pressure
 skin, 79, 142, 218
 blood, 27, 163, 164
Presynaptic, 12, 13, 18, 23, 84–87, 184, 203
Primary auditory cortex (A1). *See* Auditory cortex
Primary visual cortex (V1)
 anatomy of, 20, 31, 33–36, 70, 73, 136–142, 169, 170
 inputs to, 18, 36
 neural activity, 38, 40–42, 71, 198–201, 226, 244, 256–262, 320
 outputs from, 29, 31, 91, 136
Private, 49, 278, 302–304
Probability, statistical, 106, 121, 122
Projectional nature of representation, 45, 217, 222, 235, 299, 322
Propofol, 189, 192, 194
Proposition, propositional, 67, 117, 256, 287
Proprioceptive, 31, 33, 50, 51, 54, 135, 144, 235, 239, 299
Proprioceptor, 79, 88, 213
Prosopagnosia, 63, 74, 76
Prototype, 95–98
Psychofunctionalism, 115–118, 227
Public language, 301, 302
Pulvinar. *See* Thalamus, pulvinar
Purkinje cell, 13, 151–155, 168
Putamen. *See* Striatum
Pyramidal cell, 12, 14, 17, 20–22, 91, 96, 136, 138, 145, 146, 203

Qualia, 49, 56, 81, 101, 102, 217, 220, 291, 298–300
Qualities
 perceptual, 56, 291, 304 (*see also* Qualia)
 primary, 101, 102
 qualitative sensation, 101, 102
 secondary, 101, 102, 322
 sensory, 48, 101, 102, 214, 218
Quantum mechanics, 127, 292, 304
Quasi-stability, 216, 249, 307

Rapid eye movement (REM). *See* Sleep
Reactivation. *See* Replay
Read out, 38, 111, 217, 262, 265–268, 274, 276. *See also* Decode
Receptive field, 31, 32, 39, 142–144, 201, 226, 237, 243
Recognition, 32, 73, 98, 223, 227, 254, 258, 259, 299
 face, 62, 63, 75, 76, 90
 object, 74, 227, 240
 pattern, 94, 95, 100
Recollection, 23, 58, 68, 212
Reconstruct, reconstruction, 43, 76, 113, 288, 310
Recurrency, 87, 168, 169, 172, 199, 200
Recurrent network, 86, 87, 89, 92–95, 98, 100

Recursion, 169, 172
Reductionism, reducibility, 129, 289, 290, 292, 295
Reference, external, 182, 283
Reference frame. *See* Frame of reference
Reflex, reflexive, 7, 22, 25, 27, 28, 54, 55, 64, 65, 134, 135, 155, 166, 293
Reinforcement learning, 54, 86, 89–91, 120, 162
Reinforcing feedback. *See* Reinforcement learning
Relational, 218, 246, 247, 248, 252, 262, 270, 305, 311
Replay, 173, 181, 184–190, 207, 242
Representation
 abstract, 254, 255, 303
 arbitrary, 247, 266, 268, 277, 278, 284
 concept of, 2, 10, 42, 43
 contextual, 132, 279
 explicit, 99, 255, 274
 function of, representational function, 131–135, 167, 210, 270, 288, 309
 goal, 64
 motor, 261
 multimodal, 132–135, 160, 167, 168, 189, 193, 210, 231, 281, 300, 306–309
 representational levels, 222, 265–273, 276, 277, 280–284, 288–294, 306
 representational problem, 9, 132, 134, 163, 210
 representational power, 64, 273, 281, 320
 representational complexity, 281, 306, 312, 321
 richness (*see* Consciousness)
 of situation, 132–135, 167–171, 231, 260, 278, 306, 309, 322
 semantic, 261
 spatial, 132, 234, 237, 261
 situational (*see* Representation of situation)
 symbolic, 117, 271, 273, 277 (*see also* Symbol)
Representationalism, 295, 297–301, 303, 315, 320
Reptile, 29, 30, 309
Reticular nucleus. *See* Thalamus
Retina, 18, 36, 43, 44, 50, 71, 91, 102–105, 153–155, 165, 195, 292, 302
Retinotopic. *See* Map
Reverberatory, reverberation, 111, 166, 169, 181
Reverse psychology, 310, 312
Reward value. *See* Emotion, emotional value
Rhythm, rhythmic, 19, 21, 28, 138, 166, 173–183, 186, 200–204, 280. *See also* Oscillation
 alpha, beta, delta, gamma, theta rhythm (*see* Oscillation)
Ripple, 176, 178, 179, 181, 185–188
Robot, 54, 112, 212, 282, 283, 287, 310–312
Robustness, 98, 216, 241, 245, 262, 270
Rock, rocky network, 110–114, 129
Rotating snakes illusion, 226, 298
Rubik's Cube, 223, 224, 259, 271

Saccade. *See* Eye
Scalar, 88, 167

- Scene**
 auditory, 79, 309
 construction, 113, 125, 280
 perception of a scene, 64, 121, 222, 230
 segmentation, 242
 visual, 36, 45, 50, 51, 102, 132, 182, 188, 242, 310, 312
- Schizophrenia**, 149, 159, 182, 229, 230, 300, 313, 314
- Scotoma**, 70, 150
- Searle, J.**, 107, 108, 113, 291, 297–301
- Sea squirt, *Ascidiaeae***, 64
- Sejnowski, T.**, 100, 104
- Self**
 as agent, 56, 228, 230
 notion of the self, 8, 56, 228–230, 235, 239, 299
 self as own body, 235, 239
 self-awareness (*see* self-consciousness)
 self-consciousness, 8, 179, 228–231, 299
 self-generated, 135, 300, 311
 self-identification, 220, 221, 247, 263 (*see also* Modality)
 self-localization, 33, 299
 self-organize, self-organization, 90, 98, 99, 113, 181, 189, 190, 195, 278, 280
 self-similarity, 41, 169
- Semantic**
 content, 114, 123, 127, 291, 296
 dementia, 58, 253, 259
 knowledge, 255, 259, 260, 262
 memory (*see* Memory)
 module or system, 258–262, 273
 network, 114–117, 259
 representation (*see* Representation)
- Sense**
 of balance, 64, 133
 impression (*see* Impression)
 of presence, 214
 sense data theory, 297, 298, 301
 of space, 229, 235
 of temporal continuity, 51, 52, 59
 of time, 164, 182, 212, 229
- Sensory**
 dimension, 8, 48, 49, 88, 89, 102, 167, 223, 270, 281
 integration (*see* Integration)
 quality (*see* Quality)
- Serotonin**, 14, 28, 176, 178
- Set**, 204, 207, 319
- Sexuality, sexual**, 29, 35, 60, 163, 164, 182
- Shakespeare, W.**, 93, 95–98
- Shannon, C. E.**, 121, 123, 127, 296
- Shape**
 tactile, 79, 144, 259
 visual, 48, 50, 70, 74, 88, 197, 240, 248, 249, 259
- Sharp wave, sharp-wave ripple (SWR)**. *See* Ripple
- Shoemaker, S.**, 233, 298
- Simulation, internal**, 186, 312
- Simultanagnosia**, 242
- Singer, W.**, 200–202
- Singularity, perceptual**, 219, 220
- Situation, situational**. *See also* Representation
 multimodal, 247, 248, 281
 sketch, 210, 281, 306
 update, 135, 217
- Skill**, 28, 31, 58, 66, 182
- Sleep**
 deep, 19, 28, 174, 181, 189, 190, 210 (*see also* slow-wave; non-REM)
 and hypothalamus, 29, 164
 non-REM, 114, 173–178, 180–185, 188, 189, 318
 nuclei controlling, 28
 rapid eye movement (REM), 173–185, 188
 replay. *See* Replay
 sleeping sickness. *See* Encephalitis
 sleep-wake states, 29, 114, 125, 173
 slow-wave, 126, 128, 174–176, 179–181, 190, 318 (*see also* non-REM)
 stages (*see* non-REM; rapid eye movement; deep)
 Slow wave, slow rhythm. *See* Sleep, slow-wave
- Smell**, 55, 62, 76, 78, 101, 102, 125, 146–149, 214, 218–221, 281, 282, 307
- Sodium**. *See* Na^+ channel
- Solipsism**, 301, 303, 304
- Soma**. *See* Cell body
- Somatoparaphrenia**, 229
- Somatosensory cortex**
 body map (*see* Map, body)
 primary (SI), 36, 39, 42, 54, 79, 80, 142–144, 190, 203, 205, 206, 235, 245, 250, 293
 secondary (SII), 31, 39, 79, 80, 142, 144
- Sound**
 source, 38, 235, 238
 source localization, 145, 146, 214, 235, 238
- Sparseness**, 183, 254, 260, 271
- Spatial**
 learning (*see* Learning, spatial)
 position, 39, 125, 171, 187, 214, 270
- Spectral inversion argument**. *See* Inverted spectrum argument
- Speech**, 32, 66, 78, 87, 145, 149, 159, 247
- Spike**
 phase, 184, 201, 245, 246, 251, 262, 270
 threshold, 15, 17, 85, 137, 180
 timing, 203, 244, 245
- Spinal cord**, 11, 26, 27, 30, 79, 107, 151, 214, 215, 289, 293
- Spindle**, 175, 176, 187, 188
- Spine**, 13, 21, 23
- Spinoza, Benedict de**, 7, 11, 281
- Spiny stellate cell**. *See* Stellate neuron
- Split-brain patient**, 67
- Spontaneity**, 35, 149, 159, 160, 228
- Stability**
 quasi-stable (*see* Quasi-stability)
 perceptual, 99, 215–217, 223, 260–263, 306

- State space, 98, 126
Statistical
dependence, 123–126, 163, 172, 166, 296
properties of input, 106, 111, 112, 195, 218
relationships, 90, 123, 127
Stellate neuron, 21, 136, 139, 143, 145, 147, 169
Stereodisparity, 50, 102, 227, 318
Striatum, 29–31, 61, 66, 157–163, 166, 168, 181, 184, 187, 188, 201, 204, 270
Stroke
tactile, 35, 142, 213
vascular, 48, 66, 214
Subjectivity, 5, 64, 65, 277
Submodality. *See* Modality
Substantia nigra
pars compacta, 31, 157, 158
pars reticulata, 31, 157, 158, 161
Subthalamic nucleus, 31, 157, 158, 162, 191, 192
Summary
and consciousness, 132, 133, 259
of situation, 259
Summation, 17, 18, 85, 203
Superior colliculus, 30, 39, 136, 161, 309
Superior temporal cortex, 66, 76, 171
Supervenience, 272, 281–284, 321
Suprachiasmatic nucleus, 18, 29, 164–167, 318
Supraoptic nucleus, 29, 165–167
Switch
between attractor states (*see* Attractor)
in categorization, 262
perceptual, 94, 121, 216, 223, 253
Symbol, 36, 38, 107, 108, 115–117, 265, 273–277, 284, 291
grounding problem, 117, 266, 274
Synaptic
plasticity, 12, 22, 23, 25, 83, 95, 112, 203, 268
strength, 22–24, 84, 86, 89, 92, 95, 111
weight (*see* strength)
Synchrony, 183, 184, 190, 201–204, 241–244, 249, 270, 280, 284, 319
Syntax, 114, 123, 127, 282, 291

Tacit knowledge, 53
Taste, 35, 41, 76, 78, 105, 126, 163, 167, 218, 233
bud, 78, 88, 105
Telencephalon, 25–27, 31
Teleological, 134, 156, 166
Temperature, 29, 101, 110, 111, 164, 166, 220
Temporal compression, 182, 186, 187
Temporal lobe memory system. *See* Medial temporal lobe
Temporoparietal junction, 77, 171, 193
Texture, 79–81, 218–220
Thalamocortical, 19, 28, 35, 38, 42, 54, 138, 141, 160–163, 166, 177, 180–183, 204, 210, 227, 313.
See also Corticothalamic
- Thalamus
intralaminar nuclei, 29, 192
mediodorsal, 147, 193
midline nuclei, 29
pulvinar, 29, 136, 141, 193
reticular nucleus, 18, 19, 29, 136, 141, 177–180
thalamic relay, 18, 19, 21, 29, 30, 136, 141, 147, 180, 192
ventral posterior, 79, 142, 190
Thinking, thought, 9, 53, 67, 135, 156, 159, 179, 182, 183, 228, 230, 288, 301, 303, 312
Thirst, 167
Three-dimensional percept, 50, 79, 102, 227
Time constant, 91, 155
Timing of movement, 28, 150, 151, 162
Tone, 25, 38, 145, 190
Tononi, G., 120–127, 162, 163, 312, 313
Topology
multimodal, 219, 246, 247, 307
relationships, 219, 246, 247, 274, 275
Training, 54, 58, 88, 95, 97, 237
Transcranial magnetic stimulation (TMS), 194, 196, 197, 200, 205, 293, 300, 313
Transduction (physical, sensory), 43, 62, 64, 100–103, 277, 282
Treisman, A., 241–243
Trichromatic theory of color vision. *See* Young
Turing test, 107, 129, 311, 312
Twin, 105, 301
Twin Earth argument, 301
Two-photon imaging, 25, 190, 191
Two separate brains argument, 105

Uncertainty, 121, 122, 296, 304
Unconnected set of feature detectors, 171, 241, 245, 252
Unimodal
attractor (*see* Attractor)
brain area, 171, 271
map (*see* Map)
Unity, perspectival. *See* Perspective
Up and down states, 176, 180, 181, 188, 190, 207

V1. *See* Primary visual cortex
V2, 70–73, 138, 142
V4, V4/V4α, 31, 36, 50, 56, 70, 71, 76, 198, 248, 259, 276, 281
V5, 36, 50, 69, 70, 76, 188, 200, 226, 252, 259
Value. *See* Emotional value
Van Essen, D., 32, 73, 91, 138, 139, 141
Vector, 88, 89, 92, 106, 217
Vegetative state, 29, 67, 285, 312, 313
Verbal
behavior, 65, 66, 112, 119, 193
in general, 6, 56, 73, 256, 262, 268, 301, 302
judgment, 57, 71, 246
reporting, 119, 150, 193, 198

- Vestibular, 27, 31, 33, 135, 151, 153, 155, 213, 317
Vestibulo-ocular reflex (VOR), 153–155
Viewpoint, view. *See* Perspective
Visceral, 35, 48, 149, 167, 214. *See also* Gut
Visual anosognosia. *See* Anton-Babinski syndrome
Visual detection, 205, 206, 244
Vitalism, 107, 108, 114, 288, 297
Volition, 6, 81, 216, 228, 230
Voltage
 extracellular, 21, 174, 182
 membrane, 12, 14–19, 22, 23, 85, 180
Voluntary behavior, 7, 30, 35, 66, 159, 176
- Wall Street banker’s argument, 276, 277, 283
Wakeful, wakefulness, 112, 175, 176, 179, 180,
 189–191, 260
Watson (IBM supercomputer), 3, 310, 311
Wavelength. *See* Light
Wernicke’s area, 32, 39, 66, 78, 145
Wittgenstein, L., 49, 109, 129
 duck and rabbit (*see* Jastrow)
Working memory, 51–53, 59, 118, 119, 133, 149, 150,
 157, 203, 211, 212, 234, 317
World
 model (model of the world), 100, 104, 105, 134,
 204, 220, 227, 302
 small-world architecture, 120, 168, 170, 172
 world-referenced (*see* Allocentric)
- Young, T., 102, 103
- Zombie, 66, 217, 286–288, 297, 300, 304, 305, 314

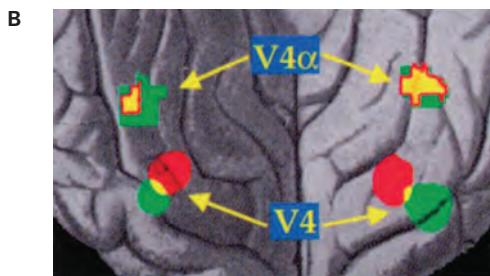
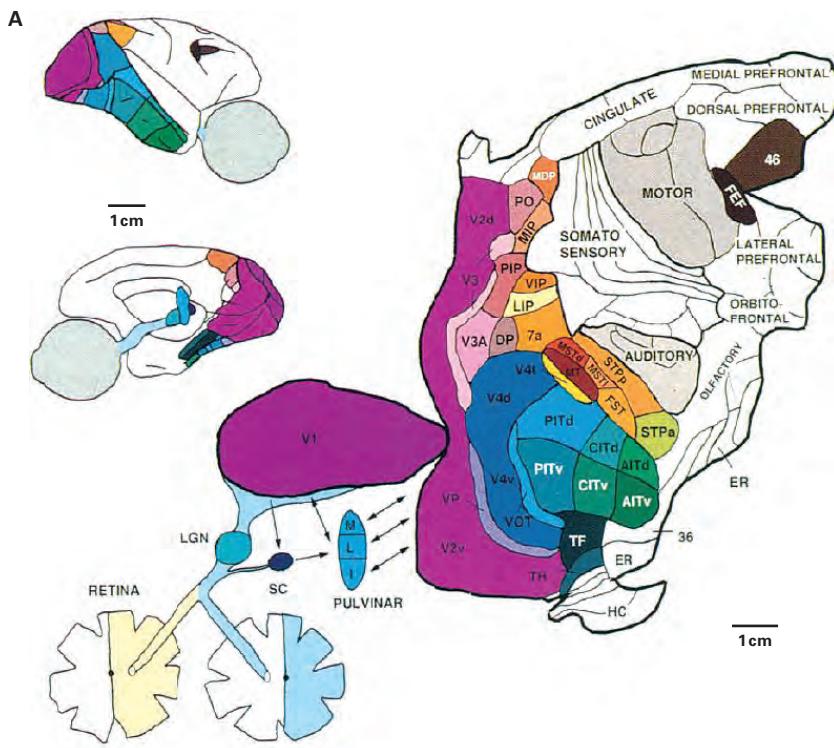


Plate 1 (Figure 3.6)

Overview of the macaque visual system and areas activated by color vision in the human brain. (A) Top and middle left insets show lateral and medial view of the right hemisphere of the macaque brain, respectively. The middle and bottom panels show the eye (gray), optic tract (light blue), lateral geniculate thalamic nucleus (LGN), superior colliculus (SC), and pulvinar as subcortical visual areas. From bottom to upper right: unfolded map of cortical visual areas shown in conjunction with subcortical structures. The three-dimensional layout captured in the left insets has been projected here on a flat surface, resulting in an artificial discontinuity between V1 and V2. From Van Essen et al. (1992), reproduced with permission from AAAS. (B) Ventral view of the V4 complex in the human brain with superimposed activation patterns derived from brain imaging. Area 4 shows evidence for a retinotopic organization, with the lower part of the visual field represented in green and the upper part in red (yellow indicates overlap between red and green). Also the more anteriorly located area V4c is activated by colored patches, set within a multicolored scene, but does not show a clear retinotopic organization. From Zeki (2005), reproduced with permission (copyright Royal Society).

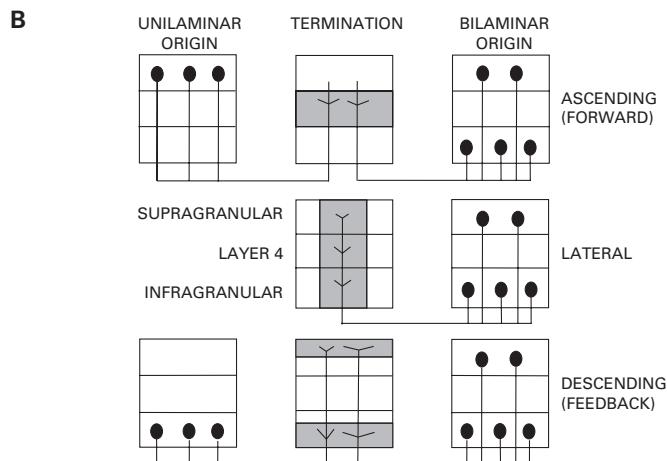
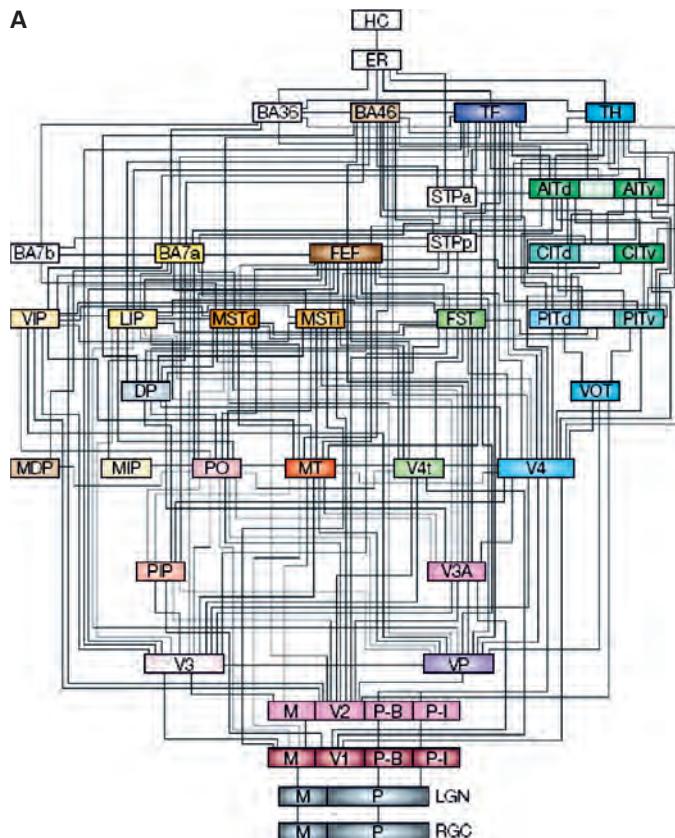


Plate 2 (Figure 6.1)

The Calling of St. Matthew by Caravaggio. This painting and its setting in the San Luigi dei Francesi church in Rome illustrate the concept of conscious experience as a dynamic, multimodal, situational representation.

Plate 3 (Figure 6.3)

(A) A hierarchy of visual areas as proposed by Felleman and Van Essen (1991), based on a compilation of anatomic tracer studies. Starting at the level of retinal ganglion cells (RGCs; bottom) and lateral geniculate nucleus (LGN; M and P denote magno- and parvocellular pathways), the projections conveying photic information spread out across a network of 32 cortical areas. A total of 187 area-area connections are shown, most of which are reciprocal. Besides up- or downward projections, many areas are connected laterally. Note how the entorhinal cortex (ER) and hippocampus (HC) involved in declarative memory are located at the apex of the hierarchy. (B) Whether an area is placed at a higher or lower level depends on its pattern of afferent and efferent connections across its laminae. All boxes shown are divided into three zones: (1) supragranular or superficial layers, (2) layer 4, and (3) infragranular or deep layers (middle panel). If a given cortical area sends a projection from its superficial layers to layer 4 of a receiving area, the projection is considered ascending or forward. The same applies if both the superficial and deep layers of the sending area are projecting. For descending projections, terminations in the superficial and deep layers of the target area are critical. Terminations of lateral connections are distributed across all three laminar compartments. Reprinted with permission from Oxford University Press.



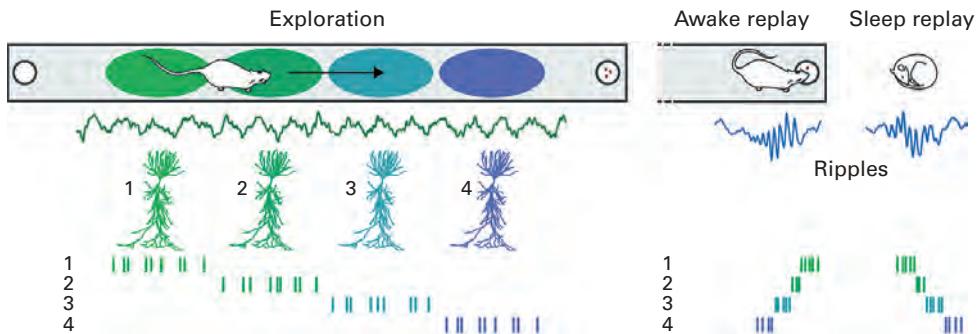


Plate 4 (Figure 7.4)

Schematic of neuronal replay associated with sharp-wave ripple activity in the hippocampus. The top left panel shows a rat running along a linear track toward a reward site at the end of the track. The trace below the track represents theta rhythm in the local field potential recorded from hippocampus during running. Below this trace, four hippocampal pyramidal neurons (numbered 1–4) are shown, each firing spikes (color coded, vertical ticks) at subsequent places on the linear track. Thus, running is associated with a sequence of place-cell firing patterns in hippocampus (lower panel). Right-hand panels: When the rat arrives at the reward site and consumes food, ripples are generated in the hippocampal EEG, during which the place-cell sequence reactivates in a temporal order opposite to that of the running sequence (reverse replay). During subsequent slow-wave sleep, place cells reactivate in the same order as during track running (forward replay). From Girardeau and Zugaro (2011), with permission from Elsevier.

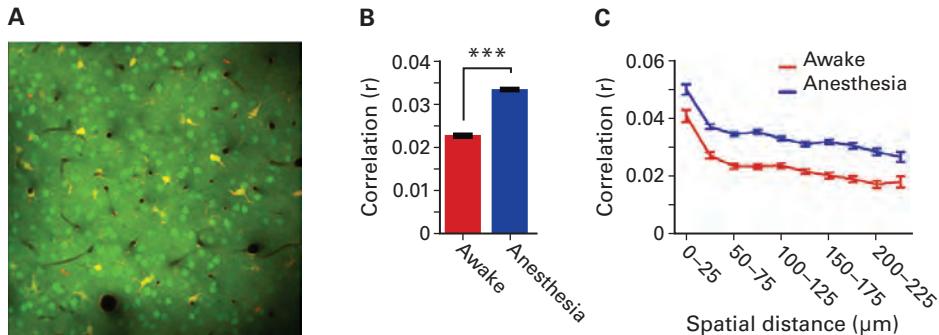


Plate 5 (Figure 7.5)

Two-photon imaging of neuronal ensemble activity in the awake and anesthetized visual cortex of the mouse. (A) Two-photon microscopy permits researchers to obtain images with sharply delineated neuronal somata (green objects) and glial cells (specifically, astrocytes; yellow objects) situated in the superficial layers of mouse visual cortex. Green fluorescent staining is achieved by loading the tissue with a Ca^{2+} -sensitive dye, Calcium Oregon Green BAPTA-AM. Spiking activity of individual neurons leads to a rise in intracellular calcium concentration, which is coupled to a transient enhancement of emitted green light. (B) When the fluorescence signals of individual cells are recorded over time, spontaneous fluctuations in electrical activity of the cells result in a pattern marked by transient peaks (firing activity) alternating with quiet periods. When cell-to-cell correlations are calculated across all pairs, studied both during wakefulness and isoflurane anesthesia, a higher correlation (shown as mean \pm s.e.m.) is obtained for the anesthetized condition (***, $P < 10^{-10}$, Wilcoxon's matched pairs signed rank test). (C) This correlation tapers off as the cortical distance between the members of a pair increases, and the difference between awake and anesthetized conditions is relatively constant across distance. From Goltstein et al. (2011); cf. Goltstein et al., 2013).

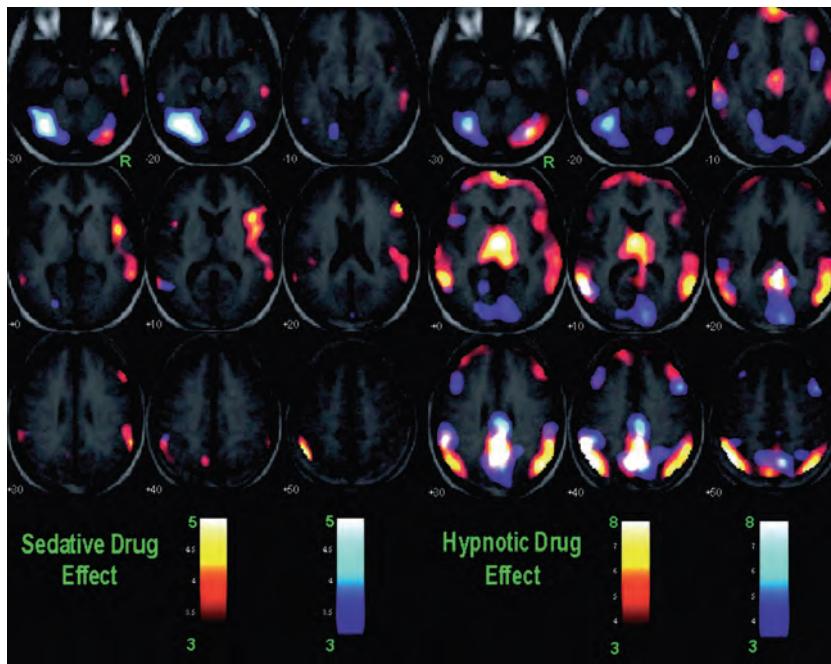


Plate 6 (Figure 7.7)

The anesthetics propofol and thiopental affect different brain regions as assessed by induced decrements in regional cerebral blood flow (rCBF). Drug effects at sedative doses (with subjects being responsive to voices or light touch) are visualized in the left-hand subset of nine brain images; drug effects at hypnotic doses (with subjects unresponsive) are on the right-hand side. Propofol and thiopental effects are rendered by a hotness (white–red) and coolness (white–blue) scale, respectively. Horizontal brain sections are shown, going from ventral (top left) toward dorsal slices (bottom right). Focusing on hypnotic effects, thiopental generally induces stronger decreases in posterior cortical areas, whereas propofol decreases activity in both posterior and more frontal regions. The two drugs also share a subset of commonly affected areas (including precuneus, cingulate cortex, inferior parietal lobule, middle frontal gyrus, middle temporal gyrus, supramarginal gyrus, and superior temporal gyrus). Some of the areas affected by hypnosis also show a decreased rCBF under sedation, when subjects are conscious but additional side effects have been documented (e.g., amnesia). From Veselis et al. (2004), with permission from Lippincott Williams Wilkins–Kluwer.

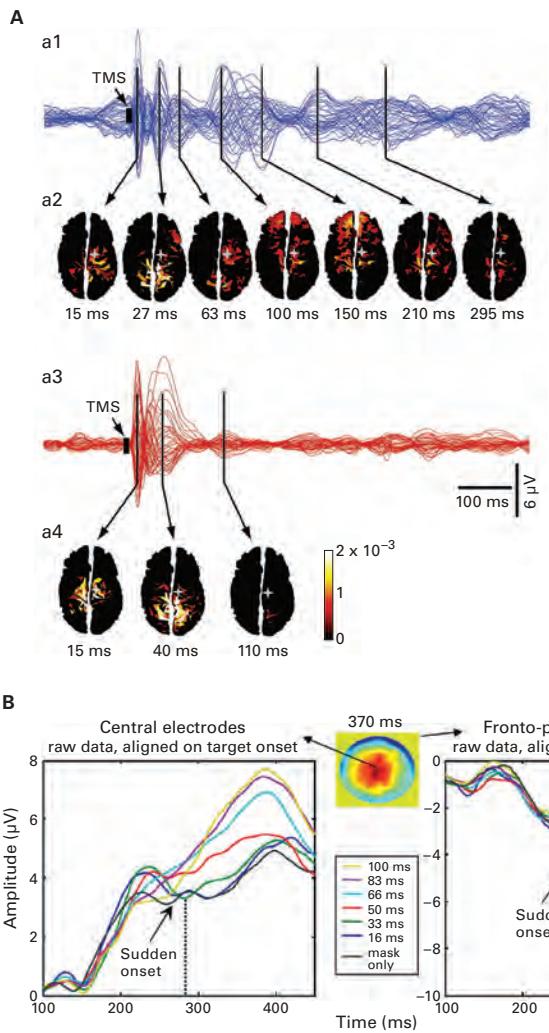


Plate 7 (Figure 7.8)

Neurophysiological studies on consciousness using anesthesia and stimulus masking paradigms. (A) Anesthesia induced by midazolam shows how scalp-EEG responses evoked by transcranial magnetic stimulation (TMS) of the premotor cortex are altered upon loss of consciousness; a1 and a3 show superimposed, averaged TMS-evoked EEG patterns for wakefulness and anesthesia, respectively. Using high-density EEG mapping, local cortical currents can be estimated (a2: wakefulness; a4: anesthesia). Wakefulness and anesthesia differ in that wakeful EEG responses are more long lasting (> 300 ms poststimulus) and occur over more distributed areas, whereas anesthesia is marked by high-amplitude, short (< 150 ms) responses that remain confined in cortical space. (B) When awake, human subjects are briefly exposed to a stimulus that is followed by a backward mask, their perception of the stimulus depends on stimulus onset asynchrony (SOA; the time interval between target stimulus and mask). In this experimental setup the threshold for reported percepts is about 50 ms. Event-related potentials (ERPs) are recorded at high density from the scalp. Left and right graphs show ERPs from central electrodes and frontopolar electrodes, respectively. Around 370 ms after stimulus onset, the central recording sites show a large peak in amplitude (left), the so-called P3 waveform. The voltage map in the middle illustrates how this waveform is spatially distributed across the scalp, using a SOA of 100 ms. When the SOA is varied from 16 to 100 ms, the central electrodes display a nonlinear increase in EEG amplitude at a latency of about 270–300 ms. The onset of this increment is more sharply aligned in time on the frontopolar electrodes. This nonlinear event can be correlated to a similar increase in subjectively reported perceptibility and has been interpreted as indicating a late “ignition” of distributed activity in fronto-parietal-temporal networks as a correlate of conscious reportability. (A) from Ferrarelli et al. (2010), with permission from the National Academy of Sciences (U.S.A.). (B) from Del Cul et al. (2007).

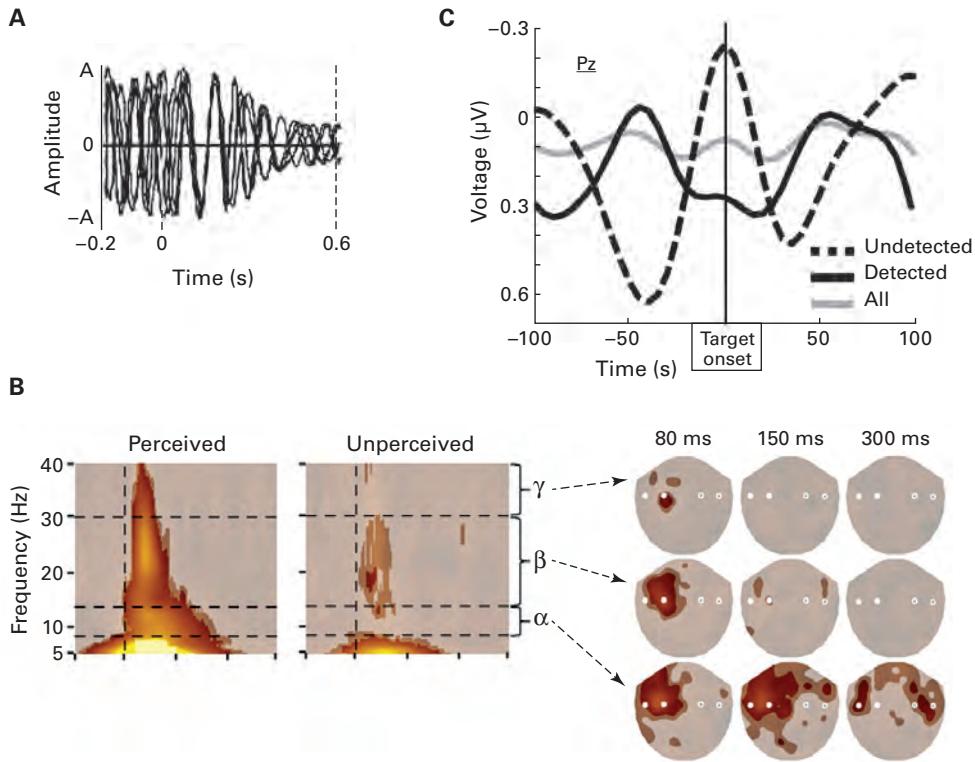


Plate 8 (Figure 7.10)

Phasing of alpha oscillations relative to perceived and unperceived stimuli. (A) Simulation of magnetoencephalographic (MEG) traces of oscillations recorded from cerebral cortex. Superimposed traces show 8–12 Hz oscillations plotted relative to onset of a weak electrical stimulus (0.2 ms) applied to the tip of an index finger. Note how ongoing alpha-band oscillations transiently phase lock to stimulus onset, that is, the oscillation has a consistent phase relationship with stimulus onset. (B) Time–frequency plots of phase locking strength over primary somatosensory cortex (contralateral to stimuli). Color transparency scales with the statistical significance of phase locking. Only consciously perceived stimuli are associated with phase locking in alpha-band oscillations. The right-hand side shows the topographic scalp distribution, across three different time points relative to stimulus onset, with predominant alpha-band phase locking across frontoparietal areas. (C) Averaged event-related potentials recorded from the midline at the level of parietal cortex (channel Pz) in a visual target detection task with metacontrast masking. Around target onset, undetected targets (dashed curve) are associated with a phase that is opposite to that of detected targets (solid curve). When detected and undetected targets are taken together, phase locking is absent (gray curve). This indicates that stimulus timing relative to the excitatory and inhibitory phases of the alpha cycle is important for conscious perception. (A, B) from Palva et al. (2005), with permission from Elsevier. (C) from Mathewson et al. (2009), with permission from the *Journal of Neuroscience*.

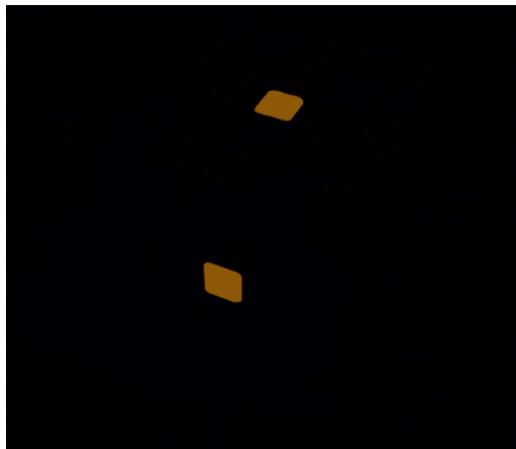
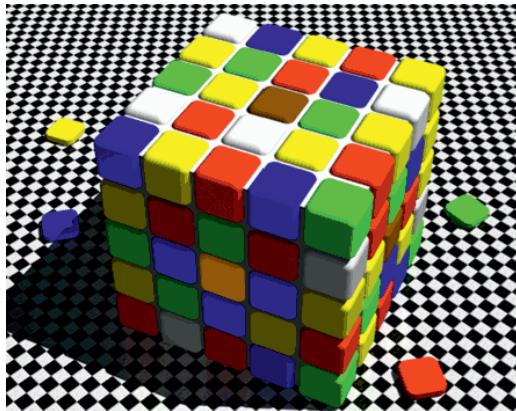


Plate 9 (Figure 8.5)

Color constancy and color illusion. Focusing on the center pieces of each of the three visible sides of this “Rubik’s Cube” in the upper panel, two are brown and the third one, on the shadow side, appears orange. In the lower panel, a dark mask has been used to remove the contextual effects of the tiles surrounding the center piece on each side. Now the colors on the top and shadow side look identical. Contextual color effects result in the same patch of color being perceived as darker when the surrounding is brightly lit as compared to the shadow side. Reproduced with permission from B. Lotto.

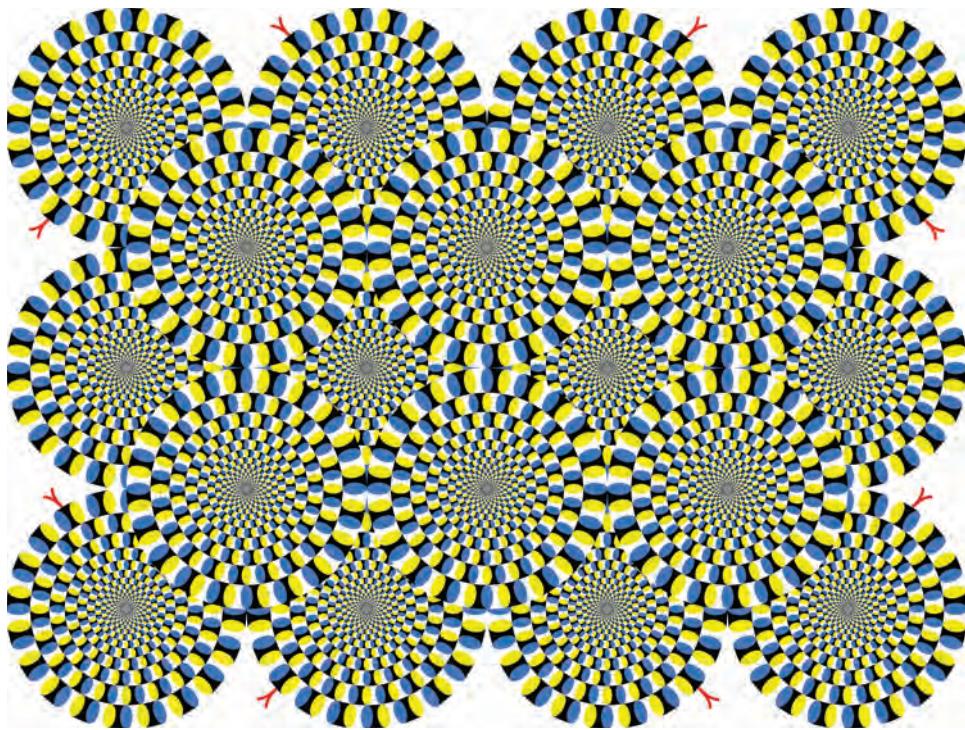


Plate 10 (Figure 8.6)

The “rotating snakes” illusion of movement. Rotations are experienced most strongly in the peripheral field of vision. Large contrast differences between adjacent segments of a coil (e.g., green–white or white–blue transitions) are important for eliciting the illusion, which might be explained from small discrepancies in processing delays in the peripheral visual system. About one in four people are not susceptible to this kind of illusion, which involves a genetic basis (Fraser & Wilcox, 1979). Reproduced with permission from A. Kitaoka, Department of Psychology, Ritsumeikan University, Japan (see also Murakami et al., 2006).

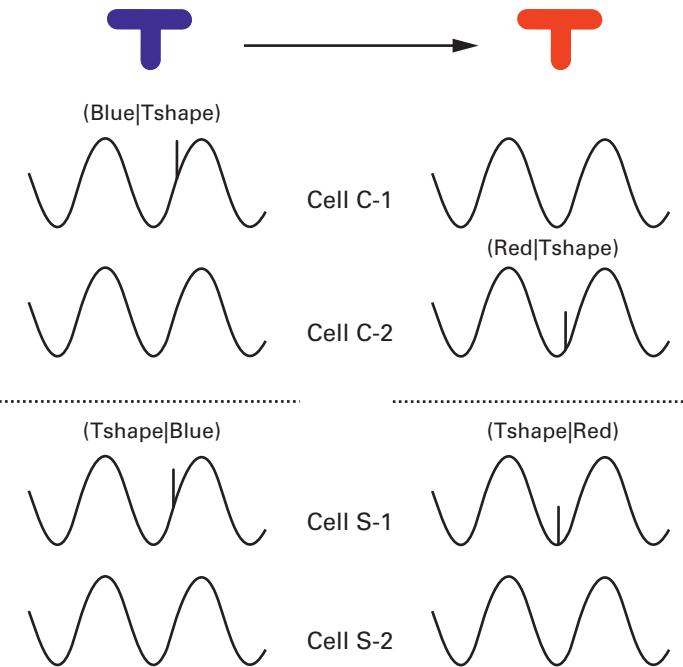


Plate 11 (Figure 9.8)

Phase coding of multimodal relationships with retention of firing-rate specificity. Upper left: The color-sensitive cell C-1 is tuned to blue and thus fires in response to blue, whereas cell C-2 is tuned to red and will remain unresponsive because the presented T-shape is colored blue. Lower left: Likewise, shape-sensitive cell S-1 is tuned to a T-form and thus fires upon presentation of the blue T-shape, whereas S-2, tuned to circles, remains silent. Upper right: When the color of the T is changed from blue to red, the red-sensitive cell C-2 will now be activated whereas C-1 will drop silent. Lower right: In the shape layer, the T-sensitive cell S-1 will remain active as the shape of the object is maintained, while S-2 remains inactive. The crucial detail is that the phase of firing of the T-responsive cell changes when the stimulus switches from blue to red while retaining the same shape. From Pennartz (2009), with permission from Elsevier.



Plate 12 (Figure 9.9)

Single-cell recordings from human epileptic patients. Recordings were made using intracranial depth electrodes implanted in brain areas such as the hippocampal formation and amygdala. The left part of the figure pertains to an amygdala neuron that responded with an increase in firing rate to pictures of animals, but not to other kinds of stimuli (e.g., faces or places; these classes are not shown). The right half is from another amygdala neuron, which fired only in response to the mouse, squirrel, and rabbit, out of 97 pictures shown. The black bar below each picture shows the time window of stimulus presentation (1 second). The display below each bar shows, for each row, the moment the cell fired relative to the stimulus, with each vertical blue tick representing one spike. Subsequent rows represent consecutive stimulus presentations (trials). Note how the right-hand neuron not only fires more selectively to the three animal pictures but also hardly fires at all to pictures of buildings and human faces. The left neuron was putatively classified as an interneuron, the right neuron as a pyramidal cell. Such cells are thought to code abstract concepts, such as “animal” or “rodent” in a sparse and explicit manner. From Quian Quiroga (2012). Reprinted by permission from Macmillan Publishers Ltd.

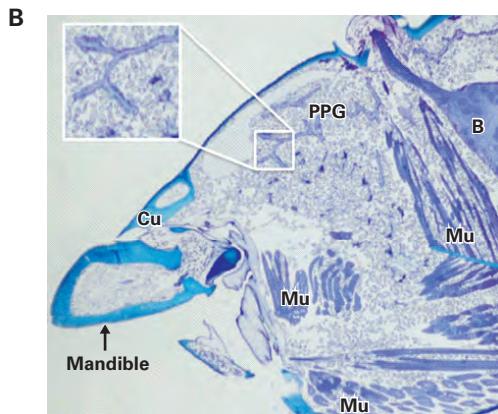


Plate 13 (Figure 11.1)

Fungus infection of the ant *Camponotus leonardi* illustrates the debate on zombies and animal consciousness. (A) Once the fungus *Ophiocordyceps* has infected the head of the ant, a stalk grows from its back to produce a fruiting body from which spores will be propulsively released. (B) Sagittal section from the head of an ant performing a “death grip” in a leaflet at the moment of fixation. Fungal hyphens are represented in the zoom-in of space around the postpharyngeal gland (PPG) and as small gray blobs filling the head and mandibles. B, brain; Mu, muscles; Cu, cuticle. The erratic random-walk behavior and the death grip of the ants can be interpreted as a form of externally induced, fungal control over the animal’s behavior, effectively reducing it to a living “zombie.” We can conceive of this behavior as happening without an “inner,” mental life, which however does not imply that mental states are separable from physical states of the nervous system. Arguments relying on conceivability or imagination suffer from the same weakness when considering human or lower-animal versions of zombies. From Hughes et al. (2011).

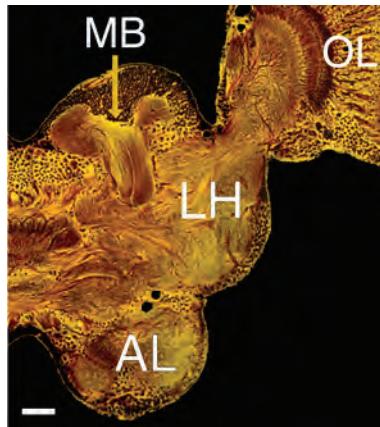
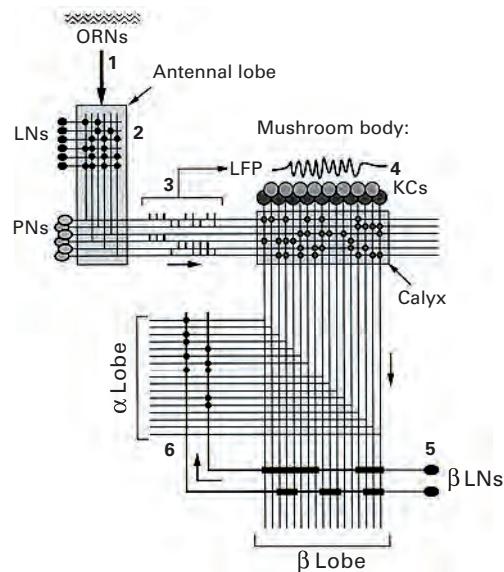
A**B**

Plate 14 (Figure 11.2)

Networks for sensory processing, multisensory integration, and memory in the locust brain. (A) Transverse section of the locust brain (Bodian stain), showing the antennal lobe (AL), mushroom body (MB), lateral horn (LH), and optic lobe (OL). Calibration: 80 μ m. (B) Chemosensory signals are relayed from olfactory receptors neurons (ORNs) on the antenna to the antennal lobe. This input is processed by projection and local neurons (PNs and LNs). Projection neurons transfer their spike signals to the mushroom body, where Kenyon cells (KCs) are rhythmically activated. Synchronized excitatory postsynaptic potentials (EPSPs) elicited by this PN input give rise to 20–30 Hz local field potential (LFP) oscillations in the calyx. In turn, Kenyon cells modify activity subsequently in the beta and alpha lobe. Several top-down feedback and multisensory input pathways are not included in this diagram (but see Hu et al., 2010, and Duistermars & Frye, 2010). The locust brain presents a puzzling case of intermediate complexity for which it is difficult to exclude the existence of basic, primordial conscious representation. It shares important functional features with corticothalamic systems of the mammalian brain, including top-down recurrent projections and multimodal integration, which are however also found in basic form in subcortical structures of the mammalian brain not implicated in conscious processing. (A) from Perez-Orive et al. (2002). Reprinted with permission from AAAS. (B) from Laurent et al. (1998). Reprinted with permission from Cold Spring Harbor Laboratory Press.