

V. Srinivasa Chakravarthy

# Demystifying the Brain

A Computational Approach



# Demystifying the Brain



V. Srinivasa Chakravarthy

# Demystifying the Brain

## A Computational Approach



Springer

V. Srinivasa Chakravarthy  
Indian Institute of Technology Madras  
Chennai, India

ISBN 978-981-13-3319-4      ISBN 978-981-13-3320-0 (eBook)  
<https://doi.org/10.1007/978-981-13-3320-0>

Library of Congress Control Number: 2018961227

© Springer Nature Singapore Pte Ltd. 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

*To,  
Sri Aurobindo and the Mother*

# Preface

“The human brain is the most complex organ in the body” “The brain is the most complex thing in the universe,” and therefore, “We won’t be able to understand the brain.” This is just a tiny bite of unqualified, unreasonable adulation that the brain receives in popular literature. There is a rather unhealthy tendency in popular media to portray the brain as some sort of a god-organ. It creates around the brain an agnostic mystique, an impenetrable aura that is only meant to be admired but never understood.

The vague and confusing explanations of brain function that are often offered by textbooks, and therefore by experts alike, do not help to dispel the mystique. For example, planning and coordination are said to be the functions of the prefrontal cortex, but cerebellum, the textbooks tell us, shares the same functions. Similarly, memory is said to be the function of both the prefrontal cortex and hippocampus. But why does the brain engage multiple systems to perform the same duty? Consider another example of an explanation that does not explain much. The thalamus, a massive portal to sensory information streaming into the brain, is called a “relay system” which means that the thalamus merely passes on the influx of signals beyond. But why does the brain need a whole complex organ to pass on incoming signals intact; a bundle of fibers would do the job. In such situations, as in a thousand others, the eager student of neuroscience is quickly told answers to a large number of questions of “what” category, but rarely “how” and almost never “why.” Such a fundamental restriction makes the brain, despite all goodwill and intent to understand on the part of an earnest student, unfathomable.

The reason behind this mysteriousness of the brain is not merely its complexity, as popular media again would like us to believe. The A380 and the International Space Station are no doubt some of the most complex systems that humans have ever created. But we are able to handle and master that complexity because we know the underlying physical principles. The complexity in the details can be effectively handled by organizing manpower or by the use of computational power. Complexity can be mastered by comprehending the principles. When we separate principles from the details, the complexity does not disappear but merely ceases to

be intimidating. The sense of jaw-dropping wonder gives way to satisfaction that comes from perfect understanding.

But when it comes to brain science, the distinction between principles and details varies from being weak to nonexistent. One wonders whether it is not just brain science, but a good part of biology that suffers from this tendency. The strongly descriptive and information-rich traditions of biology—particularly classical biology—stand in stark contrast to modern physics and engineering where the principles are primary, and the details are handled effectively and effortlessly thereof.

This near lack of discrimination between principles and details in biology has been brought to fore by molecular biologist Yuri Lazebnik in a regaling article titled: “Can a biologist fix a radio?—Or, what I learned while studying apoptosis.” Lazebnik considers a quaint thought experiment of how a biologist would proceed to understand the operation of a radio. Since biologists do not believe that physics can be of much use in their pursuit, they launch their own, unique biology-style attack on the problem of the radio. First, they would get enough funds and procure a large number of radios. They would then embark on a bold study of the radios and their constituents in gruesome detail. The vast universe of radio components—yellow and blue, spherical and cylindrical, striped and otherwise—is painstakingly mapped and embedded in an impressive taxonomy. That consummates a valorous course of structural research on the subject.

Next follows a functional study. Our biologists with their unflagging energy would now begin to pluck out components of the radio one at a time and study the effect of the missing component on the radio’s key function—to produce intelligible sounds. This new line of effort may reveal that certain components are not crucial, since when these are plucked out, the radio sputters and hisses but does not fail to make itself heard. But there are other components—perhaps a wire that connects the circuit board to the battery—in whose absence the radio is practically dead. The discovery marks a tremendous breakthrough in our biologically inspired study of the radio. It is doubtful if this line of research would consummate in a humanly meaningful time frame.

By contrast, the study of radio that is armed with a prior understanding of physical principles of the radio would proceed very differently. Basically, a radio picks up electromagnetic signals from the ambience, amplifies them, converts them into audible sounds, and plays them. Each of these steps requires a certain device, a mechanism, which can take a variety of possible physical implementations. But once we know the framework, the overall pattern, we would look for the substrates for that pattern in the physical system and quickly identify them. While a biology-style investigation may take decades to unravel a radio, an approach based on an understanding of the underlying principles, assuming they are readily available, might take a week or two, even in case of a radio of an extremely novel design.

What then is the situation in neuroscience? Do we deal today in terms of principles of brain function, or are we willingly stuck in the quicksand of details? A revolution has begun in brain science about three decades ago, though the first seeds have been sown more than half a century ago. The goal of this revolution is to

answer every possible “why” about the brain, by unearthing the principles of brain function. It has given us the right metaphor, a precise and appropriate mathematical language which can describe brain’s operations. By the application of these principles, it is now possible to make sense of the huge sea of experimental data, resolve long-standing points of confusion, and truly begin to admire the architecture of the brain. To borrow an analogy from astronomy, the new mathematics is drawing us away from the era of “epicycles,” ushering in the era of “inverse square law and Lagrangian dynamics.”

Researchers of the new computational and mathematical neuroscience have unearthed a small set of principles of *Neural Information Processing* as they are often called. As it happens in physics, researchers succeeded in explaining a wide range of neural phenomena with the same compact set of principles. That set may not be complete. There might be other principles yet to be discovered. But what has already been discovered is enough to create confidence in the existence of such a complete set. The first of these principles is the idea that information is stored in the form of strengths of connections among neurons in the brain, and learning entails appropriate modification of these connections. There are precise rules that describe such modification. Then, there is the idea that memories are stored as persistent states, the “attractors,” of brain’s dynamics or the idea that synchronized activity of neurons in distant parts of the brain has a great significance, not only to sensory-motor function, but also to more intriguing phenomena like conscious awareness. There are some more.

This book is about the neural information processing principles, since the aim of this book is to demystify and deconstruct the brain. Chapter 1 in the book, as it presents a brief history of ideas about the brain, also introduces some of the key ideas and concepts. Chapter 2 sets out to understand the logic of brain’s anatomy. It takes the reader on a quick journey through the evolutionary stages in the brain and seeks to explain some of the broad stages in that development using the minimum wire principle. Chapter 3 is an introduction to the neuron and mechanisms of a neuron’s electrical and chemical signaling. Chapter 4 takes up the neuron model just introduced and presents a simple mathematical model of the same. Using this neuronal model, Chap. 4 shows how to construct complex networks that can explain a variety of phenomena from psychology. Chapters 5 and 6, on memory and brain maps, respectively, use mathematical models to explain how memories are represented in the brain and how the formation of brain maps can be explained. Chapters 7 and 8 describe the architectures of the brain systems that process vision and touch senses, respectively. Chapter 9 is about motor function, about the brain makes life go. Chapter 10 presents a history of theories of emotions and introduces some of the key neurobiological substrates of emotion processing. Chapter 11 on language deals with the essential language circuits in the brain and describes how words are represented and produced. It does not discuss more advanced aspects of sentence-level processing. Chapter 12 takes up the conundrum of consciousness

from a neuroscience perspective. After briefly touching upon several philosophical approaches to the problem, it presents some elegant experimental approaches to this intriguing question, concluding with an outline of some of the contemporary neuroscientific theories of consciousness.

Chennai, India

V. Srinivasa Chakravarthy

## Acknowledgements

An initial form of this book was written with the kind support of the National Mission on Education through Information and Communication Technology (NME-ICT) program launched by the Ministry of Human Resource Development. I would like to express my gratitude to Prof. Mangalsundar, friend and colleague, who extended an unvarying support throughout the preparation of the manuscript. It is due to his vision and commitment that a book on popular science is included for the first time in the agenda of NME-ICT.

Sincere thanks to Prof. S. Bapiraju, University of Hyderabad; Prof. Rohit Manchanda, Indian Institute of Technology Bombay; and Prof. Srinivasa Babu, Christian Medical College, Vellore, whose meticulous and thorough reviews of the book greatly helped in perfecting the presentation.

Thanks are also due to my students Bhadra Kumar and Shruthi Krishna for helping with the artwork in the chapter on motor function; to Pragathi Priyadarshini, Nandini Priyanka, Asha Kranti, Vignan Muddapu, Anila Gundavarapu, and Dipayan Biswas for their excellent proofreading effort.

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Brain Ideas Through the Centuries . . . . .</b> | <b>1</b>  |
|          | The Beginnings . . . . .                           | 1         |
|          | Anatomy . . . . .                                  | 5         |
|          | Electrophysiology . . . . .                        | 8         |
|          | Pharmacology . . . . .                             | 10        |
|          | Clinical Studies . . . . .                         | 12        |
|          | Psychology . . . . .                               | 15        |
|          | Summary . . . . .                                  | 18        |
|          | References . . . . .                               | 19        |
| <b>2</b> | <b>Brain—Through the Aeons . . . . .</b>           | <b>21</b> |
|          | The Anatomy of Intelligence . . . . .              | 21        |
|          | The Evolution of the Nervous System . . . . .      | 25        |
|          | Hydra . . . . .                                    | 25        |
|          | Jellyfish . . . . .                                | 27        |
|          | Earthworm . . . . .                                | 29        |
|          | Octopus . . . . .                                  | 30        |
|          | Songbirds . . . . .                                | 32        |
|          | Rat Intelligence . . . . .                         | 34        |
|          | Chimpanzee Intelligence . . . . .                  | 35        |
|          | Earmarks of a Smart Brain . . . . .                | 37        |
|          | The Logic of Brain's Organization . . . . .        | 40        |
|          | Component Placement Optimization . . . . .         | 42        |
|          | Placement . . . . .                                | 44        |
|          | Routing . . . . .                                  | 44        |
|          | The Placement Problem in Neuroanatomy . . . . .    | 46        |
|          | Smart Wiring . . . . .                             | 53        |
|          | References . . . . .                               | 56        |

|          |  |     |
|----------|--|-----|
| <b>3</b> | <b>The World at the Level of a Neuron . . . . .</b>  | 57  |
|          | Vehicles of Love and War . . . . .                   | 58  |
|          | The Neuron . . . . .                                 | 62  |
|          | Electrochemistry of a Neuron . . . . .               | 65  |
|          | The Explosive Neural Response . . . . .              | 69  |
|          | The Hodgkin–Huxley Experiments . . . . .             | 72  |
|          | The Neuronal Handshake . . . . .                     | 75  |
|          | The Neuron Sums It All Up . . . . .                  | 79  |
|          | References . . . . .                                 | 81  |
| <b>4</b> | <b>Networks that Learn . . . . .</b>                 | 83  |
|          | Why Neurons are not Logic Gates . . . . .            | 83  |
|          | Perceptrons . . . . .                                | 87  |
|          | Multilayer Networks . . . . .                        | 96  |
|          | Learning Past Tense . . . . .                        | 102 |
|          | NETtalk: A Network that Can Read . . . . .           | 104 |
|          | References . . . . .                                 | 106 |
| <b>5</b> | <b>Memories and Holograms . . . . .</b>              | 109 |
|          | Shocks that Elicit Memories . . . . .                | 109 |
|          | Memories as Holograms . . . . .                      | 112 |
|          | Recurrent Networks . . . . .                         | 118 |
|          | Synapses that Memorize . . . . .                     | 126 |
|          | A Scratchpad of Memory . . . . .                     | 129 |
|          | Acetylcholine and Hippocampal Machinery . . . . .    | 133 |
|          | Sleep, Dreams, and Memory . . . . .                  | 137 |
|          | References . . . . .                                 | 139 |
| <b>6</b> | <b>Maps, Maps Everywhere . . . . .</b>               | 141 |
|          | The Self-organizing Maps . . . . .                   | 146 |
|          | Mapping the Bat’s Brain . . . . .                    | 151 |
|          | Dynamic Reorganization in Somatotopic Maps . . . . . | 154 |
|          | Where Exactly Is My Hand? . . . . .                  | 158 |
|          | Mapping the Parts of Speech . . . . .                | 163 |
|          | Discussion . . . . .                                 | 166 |
|          | References . . . . .                                 | 168 |
| <b>7</b> | <b>Pathways of Light . . . . .</b>                   | 169 |
|          | Shaping the Eye . . . . .                            | 170 |
|          | Capturing the Image . . . . .                        | 174 |
|          | The Primary Visual Cortex . . . . .                  | 187 |
|          | Visual Maps and Cortical Blindness . . . . .         | 192 |
|          | V2 . . . . .   | 195 |
|          | Perceiving Movement . . . . .                        | 197 |
|          | Recognizing Complex Objects . . . . .                | 202 |

|   |            |
|---|------------|
| Pathways of Knowing and Doing .....                 | 206        |
| Beyond the Visual Cortex .....                      | 209        |
| References .....                                    | 209        |
| <b>8 Feeling the World .....</b>                    | <b>211</b> |
| A Philosophical Touch .....                         | 213        |
| Neglected Touch .....                               | 214        |
| Touch in Human Interaction .....                    | 215        |
| Infants Need Touch .....                            | 216        |
| Touching Adults .....                               | 218        |
| The Engines of Touch .....                          | 221        |
| The Somatosensory System .....                      | 225        |
| Dermatomes .....                                    | 226        |
| The Somatosensory Cortex .....                      | 229        |
| Recognizing Objects Through Touch .....             | 231        |
| Constructing the Body Image .....                   | 236        |
| The Out-of-Body Experience and the Body Image ..... | 236        |
| References .....                                    | 241        |
| <b>9 Life in Motion .....</b>                       | <b>245</b> |
| Primeval Motion .....                               | 245        |
| Strands that Pull .....                             | 249        |
| The Innards of a Muscle .....                       | 254        |
| The Motor Unit .....                                | 256        |
| Spinal Circuits .....                               | 261        |
| Spinal Control of Locomotion .....                  | 267        |
| Motor Cortex and Willed Action .....                | 271        |
| Moving Willfully .....                              | 277        |
| References .....                                    | 283        |
| <b>10 Circuits of Emotion .....</b>                 | <b>285</b> |
| Ancient Emotions .....                              | 286        |
| Emotions in Psychology .....                        | 287        |
| The Unconscious Depths of Emotions .....            | 295        |
| Animal Emotions and Facial Expressions .....        | 299        |
| Emotions Right in the Middle .....                  | 301        |
| The Middle Kingdom of Emotions .....                | 304        |
| Almond Fears .....                                  | 306        |
| Memorizing Fear .....                               | 311        |
| Brain Mechanisms of Pleasure .....                  | 313        |
| Summary .....                                       | 317        |
| References .....                                    | 318        |

|   |     |
|---|-----|
| <b>11 A Gossamer of Words . . . . .</b>                       | 321 |
| Ascent of the Word . . . . .                                  | 327 |
| Mechanisms of Reading Words . . . . .                         | 333 |
| Understanding Dyslexia . . . . .                              | 338 |
| Language of the Hemispheres . . . . .                         | 340 |
| Other “Signs” of Language Impairment . . . . .                | 344 |
| References . . . . .  | 347 |
| <b>12 The Stuff that Minds Are Made of . . . . .</b>          | 349 |
| Seeing—Consciously or Otherwise . . . . .                     | 354 |
| On Being Aware of Being Touched . . . . .                     | 364 |
| The Subjective Timing Experiments of Benjamin Libet . . . . . | 367 |
| Distortions in Doership . . . . .                             | 370 |
| Varieties of Consciousness . . . . .                          | 374 |
| References . . . . .  | 377 |

## About the Author

**V. Srinivasa Chakravarthy** obtained his Ph.D. from The University of Texas at Austin and completed his postdoctoral training at Baylor College of Medicine, Houston. He is currently a professor of biology at the Indian Institute of Technology Madras, India. His research interests are in the areas of computational neuroscience and machine learning. In computational neuroscience, his research focuses on the modeling of basal ganglia to understand Parkinson's disease, modeling of neuron-astrocyte-vascular networks, and modeling of spatial cells of the hippocampus. Further, his work on character recognition in Indic scripts led him to develop the Bharati Script—a simple and unified script that can be used to express all major Indian languages.

# Chapter 1

## Brain Ideas Through the Centuries



*My hand moves because certain forces—electric, magnetic, or whatever ‘nerve-force’ may prove to be—are impressed on it by my brain. This nerve-force, stored in the brain, would probably be traceable, if Science were complete, to chemical forces supplied to the brain by the blood, and ultimately derived from the food I eat and the air I breathe.*

—Lewis Carroll (1832–1898), from *Sylvie and Bruno*, 1890.

### The Beginnings

The story of what the human brain thought of itself over the millennia would be a very interesting read. From the days when men were not even certain about the status of the brain as the seat of mind and intelligence, to the present times of gene therapies and deep brain stimulation, brain science has come a long way. Like any other science history, history of the brain is a history of errors in our ideas about the brain. A study of historical questions in this science, followed by an account of some of the questions answered (or remain unanswered, like the vexing question of “consciousness”) in contemporary thinking, helps us arrive at a balanced and realistic perspective of contemporary knowledge in neuroscience.

The father of Western medicine, Greek physician, Hippocrates (460–379 B.C.), believed, as we do now, that brain is responsible for sensation and is the seat of intelligence. Plato, who is known to us for his ideas of the republic, for his imaginings of an ideal society, for his memorable dialogues in philosophy, also thought of brain on similar lines. But his famous disciple, Aristotle, who held views (many of them dead wrong) on a wide variety of physical phenomena, believed that the *heart* is the seat of consciousness. Perhaps, he was guided by a common medical fact that a body can survive a dead brain but not a heart that had stopped beating.

**Fig. 1.1** Greek physician  
Galen



Among the ancient Greek scientists, substantial progress in understanding of the brain, particularly its structure, was achieved by Galen, one of the first Greek physicians (Fig. 1.1).

At Galen's time, the clinical medical practice was in a sort of a disarray. There was no sound scientific framework to guide clinical practice. While many blindly followed the Hippocratic tradition, others (like some present-day "holistic" clinics) used "healing music" and magical chants. A religious injunction of those times, that forbade the use of human cadavers for anatomical studies, seriously constrained progress. This forced Galen to study animal cadavers and extrapolate those observations to human anatomy. He mastered the art of dissection, wrote extensively and laid foundations to anatomical tradition. For example, in his book "On the brain" he gave precise instructions regarding how an ox' brain has to prepared and dissected:

When a [brain] part is suitably prepared, you will see the dura mater... Slice straight cuts on both sides of the midline down to the ventricles. ... Try immediately to examine the membrane that divides right from left ventricles [septum]. ... When you have exposed all the parts under discussion, you will have observed a third ventricle between the two anterior ventricles with a fourth beneath it. ...

Guided by prodigious anatomical studies, which earned him the title "restorer of anatomy," Galen learnt a lot about the structure of the brain. As the above excerpt indicates, he knew about the ventricles, the pia mater, and the hemispheres. He knew about the autonomous nerves that control internal organs like the heart and the

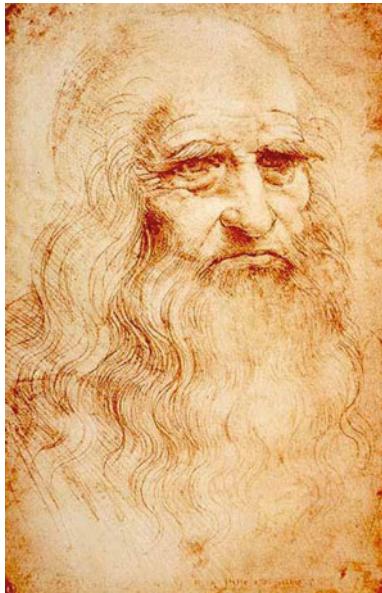
lungs. He knew of the somatic nerves that control, for example, the vocal cords. (By snapping these so-called “nerves of voice,” he demonstrated how he could silence bleating goats and barking dogs.) But when it comes to brain function, he erred deeply. A microscopic study of brain function needs a technology that would come one and a half millennia later. Thus he only speculated on brain function. He believed, just like his predecessors like Erasistrasus and others, that there exist certain “winds”—the *pneumata*—or animal spirits that surge through the hollows of the nerves and produce movement. When there are no bodily movements, these unemployed “spirits” lodge themselves in the ventricles of the brain. Thus Galen considered the ventricles to be the seat of the “rational soul.”

Galen’s case is quite representative of a line of thinking, of a puzzling dichotomy, that prevailed for nearly one and a half millennia (if not longer) in the world of neuroscience. There was a longstanding dichotomy between knowledge of structure versus knowledge of function of the brain. Those that came later in Galen’s tradition, da Vinci, Vesalius, and other great anatomists, constantly reconfirmed and expanded anatomical knowledge. But when it came to brain function, the archaic ideas of animal spirits and pneumata lived on perhaps too long. In a sense, this dichotomy in our knowledge of brain structure as opposed to that of brain function, survives even to this date. (We now have extremely detailed 3D anatomical maps of the brain, but we do not know, for example, why the Subthalamic Nucleus is the preferred target of electrical stimulation therapy for Parkinson’s disease.) The right insights and breakthroughs in our understanding of brain function, the right language and metaphor and conceptual framework, emerged all within the last half a century. These new ideas have hardly yet impacted clinical practice. We will visit these ideas, which are the essence of this book, again and again.

Leonardo da Vinci: This great artist, the creator of the immortal Monalisa, had other important sides to his personality, one of them being that of a scientist. The human cadavers that he used in his artistic study of the human figure, also formed part of his anatomical studies. His studies earned him a deep knowledge of brain’s anatomy. He likened the process of dissection of brain to the peeling of layers of an onion: to get to the brain, you must first remove the layer of hair, then remove the scalp, then the fleshy layer underneath, then the cranial vault, the dura mater... In the artist’s view, these are the brain’s onion rings. Leonardo too, like his predecessors, had knowledge of the ventricles. And like his predecessors, he erred by attributing a deep cognitive function to ventricles. He believed that the third ventricle is the place where the different forms of sensory information—sight, touch, hearing, etc.—come together. He too imagined animal spirits in the body activating limbs and producing movements. Thus, the dichotomy between knowledge of structure and function continues in Leonardo and survives him (Fig. 1.2).

Descartes: Those from the “hard” sciences know of Rene Descartes as the creator of analytic geometry, a result of the marriage of algebra and geometry. In the history of neuroscience, Descartes marks an interesting turning point. Descartes gave a new twist to the mind–body problem that has vexed all his predecessors. While knowledge of structure was founded on concrete observations, understanding of function was fantastic and often baseless. Descartes cut this Gordian knot by simply suggesting

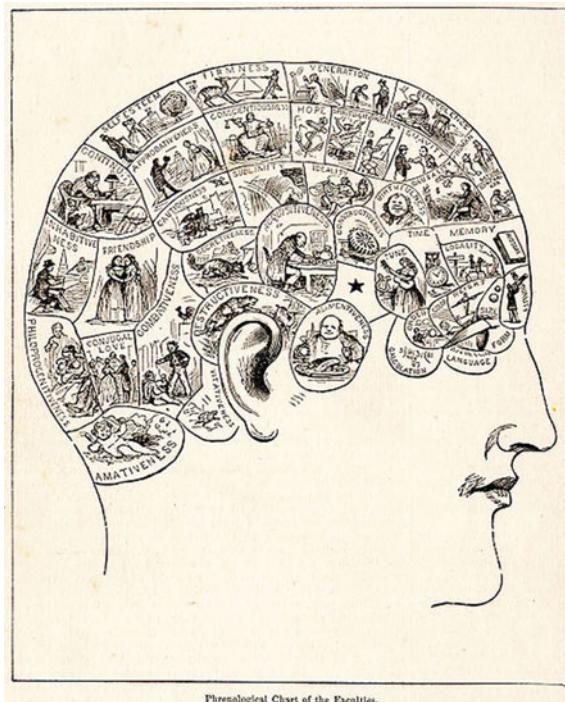
**Fig. 1.2** Leonardo da Vinci



that mind and body follow entirely different laws. Body is a machine that follows the familiar laws of physics, while mind is an independent, nonmaterial entity lacking in extension and motion. However, he allowed a bidirectional influence between the two: mind on body and vice versa. Though such pure dualism did not solve the real problem of mind versus body, it seems to have unshackled neuroscience research. It allowed researchers to ignore the “soul” for the moment, and apply known laws of physics to brain and study the “machine.” It is ironical—and perhaps has no parallel in the history of any other branch of science, that an immense progress in a field was accomplished by bypassing the most fundamental question (“What is consciousness?”) of the field and focusing on more tractable problems (e.g., “How do neurons of the visual cortex respond to color?”).

Once Descartes exorcized the “soul” from the body, it was left to the scientists to explain how the cerebral machine, or the “computational brain” in modern language, worked. Since all the cognitive abilities cannot be attributed to an undetectable soul anymore, it became necessary to find out how or which parts of the brain support various aspects of our mental life. A step in this direction was taken by a German physician named Franz Joseph Gall in 1796. Gall believed that various human qualities are localized to specific areas of the brain. This modular view of brain function is a refreshing change from the lumped model of the soul. But that’s where the virtues of the new theory end. Gall thought that the size of a specific brain region corresponding to a psychological quality is commensurate to the strength of that quality in that individual. A generous person, for example, would have a highly enlarged “generosity” area in the brain. As these brain areas, large and small, push against the constraining walls of the skull, they form bumps on the head, which can be seen

**Fig. 1.3** A map of the brain used by phrenologists



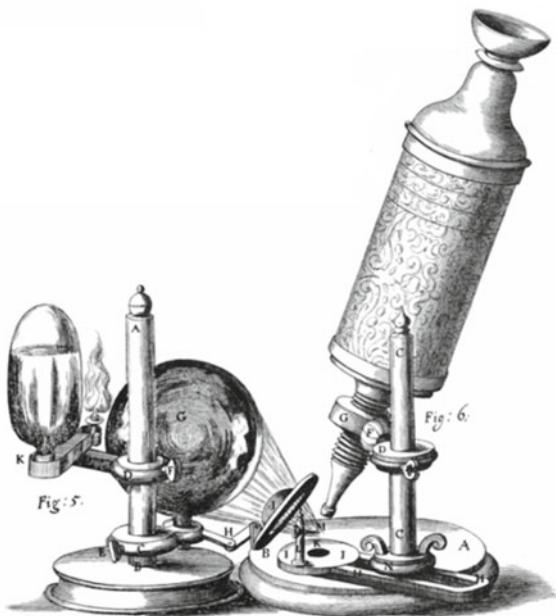
or felt by a keen observer, as the theory claimed. A person's biodata is graphically written all over the skull! This quaint science was known as Phrenology (its haters called it "bumpology"). Even in its early days, phrenology was criticized by some as a pseudoscience. Nevertheless, its followers grew and its popularity and practice survived to recent times (In 2007, the American state of Michigan began to tax phrenology services) (Fig. 1.3).

Phrenology was an interesting, though awkward, first step toward understanding localization of functions in the brain. Its strengths over the "soul theory" lie in this localization approach. But it failed to go very far since its hypotheses were not based on any sound physical theory. An ideal explanation of brain function must emerge, not out of unbridled imagination, but out of the rigorous application of physical principles to the nervous system. Thus, progress in our understanding of the brain occurred in parallel to progress in various branches of science.

## Anatomy

Knowledge of large-scale anatomy of the brain existed for at least two millennia. However, insight into the microscopic structure of the brain came with the develop-

**Fig. 1.4** The microscope used by Anton van Leeuwenhoek



ment of tools to peer into the smallest recesses of the brain. The compound microscope with illumination created by Robert Hooke gave us the first glimpses of the microstructure of the biological world. Hooke observed organisms as diverse as insects, sponges, bryozoans, or bird feathers with this new device. He made delicate drawings of what he observed and published them in the famous “Micrographia” in 1665 (Fig. 1.4).

Anton van Leeuwenhoek, who had a passion for constructing microscopes, took this tradition further, by making observations at a much smaller scale. In 1683, one day, as he was observing his own sputum in the microscope, he noted that “in the said matter, there were many very little animalcules, very prettily a-moving.” These “animalcules,” these minuscule “animals,” that Leeuwenhoek saw were the first biological cells ever observed. Subsequently, he also observed a nerve fiber in cross section.

Microscopic observations of nerve cells posed a new problem that did not exist in other tissues of the body. Nervous tissue everywhere had these long fibers connected to cell bodies. These did not resemble the blob-like cells of other tissues. It was not clear if neural tissue had discrete cells with clear boundaries separating cells. Thus, early microscopic observations led people to believe that cells in the nervous tissue are all connected to form a continuous, unbroken network—not unlike a mass of noodles—known as the “syncitium.” The limitations of early microscopy, compounded with the transparent appearance of cells, were at the root of this difficulty. It was not too long, before Camillo Golgi developed a way of “coloring” the cell, so that they stood out stark against a featureless background. Putting this Golgi staining

**Fig. 1.5** A drawing by Ramon y Cajal of a Purkinje cell, a neuron located in the cerebellum



technique to brilliant use, Ramon y Cajal observed neural tissue from various parts of the brain. Figure 1.5 shows an intricate drawing made by Cajal of Purkinje cell, a type of cell found in cerebellum, a large prominent structure located at the back of the brain.

From his observations, Cajal decided that neural tissue is not a featureless neural goo, and that it is constituted by discrete cells—the neurons. What distinguishes these brain cells from cells of other tissues are the hairy structures that extend in all directions. Cajal taught that these discrete, individualized cells contact each other using these “wire” structures. Thus, the interior of one cell is not connected to the interior of another by some sort of a direct corridor. At the point where one cell contacts another, there must be a gap. (Interestingly, the gap between two contacting neurons was too small to be observable in microscopes of Cajal’s day. But Cajal guessed right.) Thus he viewed the brain as a complex, delicate network of neurons, a view known as the “neuron doctrine.” In honor of the breakthroughs they achieved in micro-neuroanatomy, Golgi and Cajal shared a Nobel prize in 1906. Subsequently, Ross Harrison performed microscopic observations on the developing brain in an embryo. Neuron-to-neuron contacts would not have matured in the embryonic brain.

In this stage, neurons send out their projections, like tentacles, to make contact with their ultimate targets. Harrison caught them in the act and found that there exists indeed a gap, as Cajal predicted, between neurons that are yet to make contact with each other, like a pair of hands extended for a handshake.

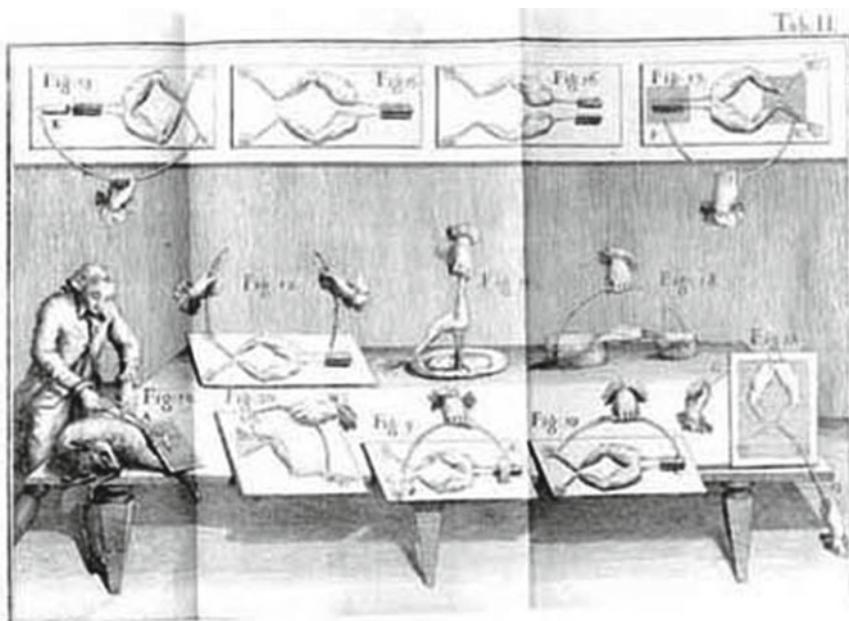
These early microanatomical studies of the brain revealed that the brain consists of cells called neurons with complex hairy extensions with which they make contact with each other. Thus brain emerged as a massive network, a feature that distinguishes itself from nearly every other form of tissue, a feature that perhaps is responsible to its unparalleled information processing functions.

Learning about brain's microanatomical structure is the first step in learning what makes brain so special. But in order to understand brain's information processing function, one must study what the neurons *do*. What is the nature of the "information" that they process? How do they produce and exchange that information? A beginning of an answer to these questions came with the realization that neurons are electrically active, like tiny electronic circuits. Progress in this line of the study came with the development of a branch of biology known as electrophysiology, which deals with the electrical nature of biological matter.

## Electrophysiology

Though classical biology teaches that all life is chemical, and solely chemical, it is equally valid to say that all life is electrical. The field of bioelectricity sprang to life on one fine day in 1771, when Italian physician Luigi Galvani observed that muscles of a dead frog suddenly contracted when brought into contact with an electric spark. When Galvani's assistant touched the sciatic nerve of the frog with a metal scalpel which had some residual electric charge, they saw sparks fly and the leg of the dead frog kick. At about that time, Galvani's associate Alessandro Volta developed the so-called Voltaic pile, which is the earliest battery or an electrochemical cell. While Galvani believed that the form of electricity found in the muscle is different from what is found in an electrochemical cell, Volta believed the opposite. Volta was right. Thus began the realization that what activates the muscle is not some mysterious "animal electricity," but the very same electricity found in a nonliving entity like the electrochemical cell (Fig. 1.6).

In the early nineteenth century, German physiologist Johannes Muller worked on the mechanism of sensation. He found that the sensation that results on stimulation of a sensory nerve depends, not on the nature of the stimulus (light, sound, etc.), but merely on the choice of the nerve. Thus when the retina, which contains a layer of photoreceptors in the eye, or the optic nerve, which carries visual information to the brain, are activated by light or pressure or other mechanical stimulation, a visual sensation follows. (This fact can be verified by simply rubbing on your closed eyes with your palms.) Muller termed this the *law of specific energies* of sensation. Muller began a tradition in which physical principles are applied without hesitation



**Fig. 1.6** Drawings by Galvani depicting his experiments with electrical activation of frog legs

to understand the electrical nature of the nervous system. In a volume titled *Elements of Physiology*, he states this perspective, though with some caution, as follows:

Though there appears to be something in the phenomena of living beings which cannot be explained by ordinary mechanical, physical or chemical laws, much may be so explained, and we may without fear push these explanations as far as we can, so long as we keep to the solid ground of observation and experiment.

Two of Muller's illustrious disciples—Emil du bois-Reymond and Hermann von Helmholtz—developed Muller's vision. Du-bois Reymond, who proceeded along experimental lines, began his career with the study of “electric fishes,” creatures like the electric eel, catfish, and others that are capable of producing electric fields. He worked extensively on electrical phenomena related to animal nervous systems and described his findings in the work *Researches on Animal Electricity*. His important contribution to electrophysiology was the discovery of the action potential, a characteristic, well-formed voltage wave that is seen to propagate along nerve fibers. But he did not possess the requisite theoretical prowess to understand the physics of the action potential.

Another event that greatly helped our understanding of the electrical nature of the brain, is a revolutionary development in our understanding of electricity itself. It took the genius of James Clerk Maxwell, the theoretical physicist who integrated electricity and magnetism in a single mathematical framework. Out of this framework emerged the idea that light is an electromagnetic wave propagating through vacuum.

Such developments in physics enabled theoretical physicists like Herman von Helmholtz to carry over these theoretical gains into the study of biology, particularly that of the nervous system. With a strong foundation in both theoretical physics and physiology, Helmholtz would have been a great presence in the interdisciplinary biology research of his times. In present conditions, typically, an individual is either an expert in biology and learns to apply ideas and core results from physics or mathematics, or an expert in physics who is trying to solve a biological problem that is already well formulated as a problem in physics. Or on occasions, a biologist and a physicist come together to apply their understanding to a deep problem in biology. But that was not to be the case with Helmholtz. Alone he progressed in both physics and biology and made fundamental contributions to either fields. Drawing inspiration from the work of Sadi Carnot, James Joule, and others, he intuited that heat, light, electricity, and magnetism represent various, interchangeable forms of energy. In thermodynamics, along with William Rankine, he popularized the notion of the heat death of the universe, which refers to the theoretical possibility that when the universe evolves to a state of maximum entropy, there will be no more free energy left to support life. He devised the so-called Helmholtz resonator that has valuable applications in acoustics. His invention of ophthalmoscope, the device that enables observation in the interior of the eye, revolutionized ophthalmology. His contribution to electromagnetism is epitomized in the famous Helmholtz equation, which describes the propagation of electromagnetic waves under special conditions. These wave propagation studies paved the way to a physics-based understanding of the propagation of action potential along the nerve.

Thus by the beginning of the twentieth century, it became abundantly clear that the brain is an electrical engine, a massive electrical circuit with neurons as circuit elements and the nerve fibers as wire. The signals using which neuron converse with each other are not too different from telegraphic signals. The “animal spirits” and “vital forces” that haunted brain science for millennia were thoroughly exorcized. Furthermore, developments in electrophysiology and electrical engineering created a sound framework for a systematic study of brain function.

## Pharmacology

But then all events in neural signaling are not electrical. When a neuron A sends a signal to neuron B, there is an important step in this process that is chemical. Neuron A releases a chemical C, which crosses the minute gap (which Cajal guessed but could not see) that separates the two neurons, and acts on neuron B. This process of neural interaction by transmission of chemicals is known as neurotransmission. A microscopic understanding of this process came about only in the later half of the past century.

However, knowledge of chemicals that act on the nervous system is perhaps as old as humanity itself. Substances that soothe or induce sleep, substances that reduce

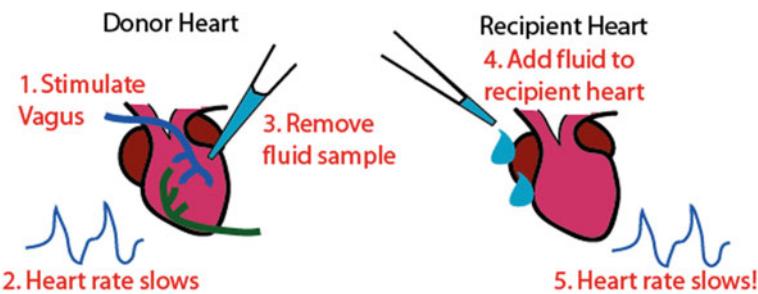
pain, poisons used by hunters to immobilize their prey without killing them are all instances of knowledge of chemicals that act on the nervous system.

In more recent history, in the middle of the nineteenth century, pioneering work on the action of drugs on the nervous system was performed by Claude Bernard, a French physiologist. Claude Bernard was known foremost for his idea of homeostasis, which postulates that the internal state of the body is maintained under constant conditions, in face of changing external conditions. In his own words, this idea may be stated as: “La fixité du milieu intérieur est la condition d'une vie libre et indépendante” (“*The constancy of the internal environment is the condition for a free and independent life*”). He studied the physiological action of poisons, particularly two of them: curare and carbon monoxide. Curare is a muscle poison traditionally used by hunters in South America. When an animal is struck by arrows dipped in curare, it dies of asphyxiation since the poison deactivates muscles involved in respiration. Carbon monoxide is a poisonous gas that acts on the nervous system producing symptoms like confusion, disorientation, or seizures. In large doses, it can cause death by destroying the oxygen-carrying capacity of the blood.

If poisons can act on the nervous system and produce harmful effects, drugs can produce therapeutic effects. This latter phenomenon preoccupied John Langley a Cambridge physiologist who studied the effects of drugs on living tissue. In the second half of the nineteenth century, the action of drugs like morphine (a sedative) and digitalis (increases cardiac contractility and counteracts arrhythmias) was explained vaguely in terms of special, inexplicable affinities of tissues to drugs. It was thought that drugs somehow directly act on the tissue/cell itself. But Langley believed that drug action is no different from chemical interaction between two molecules. Through a series of brilliant experiments, he gathered evidence to the idea that drugs act on tissue indirectly via the agency of a “receiving molecule”—which he called a receptor—that receives the action of the drug and transfers it to the surrounding tissue.

But what is the purpose of these receptors on neurons? Obviously, it cannot be that they are waiting only to bind to a drug molecule inserted by a curious pharmacologist. It then leaves the possibility that neurons talk to each other by chemicals, and receptors are the means by which a neuron understands the molecular signal transmitted by another neuron. This was the line of thought of Otto Loewi who was searching for a way of proving that neurons conversed by exchange of chemicals. Before Otto Loewi's time, it was not clear if neurons communicated electrically or chemically. Loewi devised an ingenious experiment to settle this issue. He claims that he saw the plan of the experiment in a dream.

The experiment consists of a preparation of two frog hearts. The hearts are kept alive and beating in separate beakers containing Ringer's solution. One of the hearts has the intact vagus nerve connected, which when stimulated is known to slow down the heart. Loewi electrically activated the vagus nerve, which slowed down the corresponding heart. He then took some liquid bathing this heart and transferred to the second beaker, which contained another heart. The second heart immediately slowed down. The only reasonable explanation for this effect is as follows. When the vagus nerve was activated, it released a substance which dissolved in the surrounding liquid.



**Fig. 1.7** A schematic of Otto Loewi's experiment

It was this substance that slowed down the first heart. When this liquid was transferred to the second beaker, it slowed the second heart too. Thus, the (vagus) nerve acted on the heart not by direct electrical action but by a chemical means. Subsequently, it was discovered that neurons communicated by the release of a chemical called a neurotransmitter, which is recognized by a receptor located on a target neuron. The site of this chemical exchange is a small structure—the synapse—which acts as a junction between two neurons. There were also neurons that communicated directly via electrical signaling. Otto Loewi shared the Nobel Prize with Henry Dale, who did pioneering work on acetylcholine, an important neurotransmitter. John Eccles was awarded a Nobel Prize for his work on electrical synapses. These pioneering studies became the key edifices of the vast realm of neurochemistry and neuropharmacology (Fig. 1.7).

## Clinical Studies

While studies on electrochemistry and neurochemistry revealed neuronal signaling events at a microscopic level, a wealth of information emerged from studies of patients with neurological disease. These studies allowed researchers to take a peep into how different brain regions worked together in controlling a person's behavior.

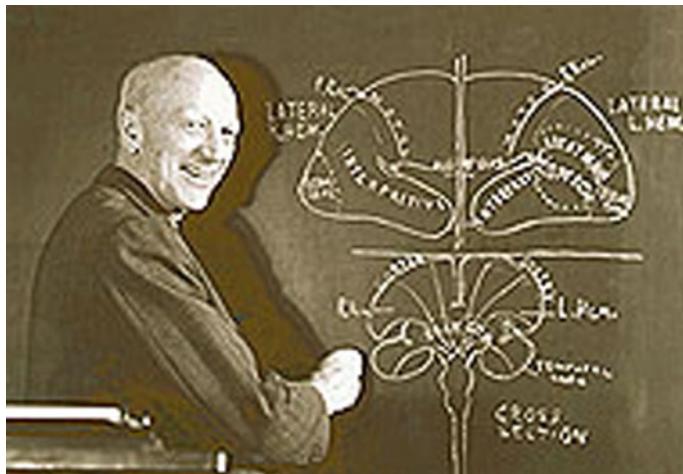
Since the time of Franz Gall and his phrenology, two rival theories existed regarding how different brain functions are mapped onto different brain structures. One theory, known as localization view, believed that specific brain structures operate as sites for specific brain functions. The contrary theory, known as the aggregate field view, claimed that all brain structures contribute to all aspects of human behavior. Phrenology itself was perhaps the first example of an extreme localization approach. However, we have seen that it is only a belief system, tantamount to superstition—without any scientific support. In the nineteenth century, a French physiologist named Pierre Flourens decided to put the localization approach to test. He took experimental animals, made gashes on their brains at various locations, and observed their laboratory behavior. He noted that changes in behavior depended not exactly on the sites of

these gashes, but only to the extent of the damage. This set of studies seemed to support the aggregate field view. Similar observations were echoed much later in the early twentieth century by Karl Lashley, who studied experimental rats engaged in maze learning. But some clinical studies told a contrary story.

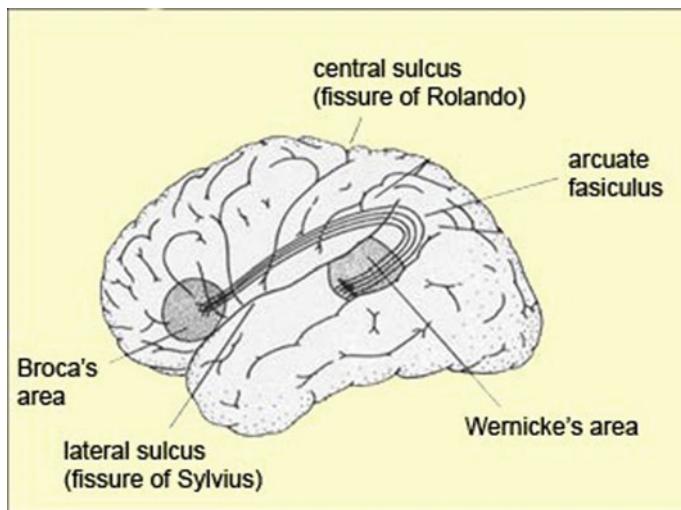
British neurologist Hughlings Jackson studied a form of seizures known as focal motor seizures, a type of uncontrollable convulsions that begin at an extremity and spread sometimes to the entire side of the body. Jackson, after whom these seizures were later named, speculated that these convulsions were probably driven by neural electrical activity that spreads, like a forest fire, over the brain surface. Since the seizures spread from an extremity to more central parts of the body, Jackson inferred that specific brain regions when activated produced movements in specific body parts. Such precise correspondence between brain regions and movements was later confirmed by electrical stimulation experiments performed by Wilder Penfield a Canadian neurosurgeon. These clinical studies seem to support the localization view (Fig. 1.8).

More support came from patients with aphasias, a general term indicating speech impairment. In one form of aphasia, named as Broca's aphasia after its discoverer, the patient has difficulty forming complete sentences, and has non-fluent and effortful speech. Utterances usually have only content words (nouns, adjectives, etc) omitting most function words (verbs, pronouns, prepositions, etc). For example, a patient who wanted to say that he has a smart son who goes to university might end up saying something like: "Son...university... smart...boy." In extreme cases, a patient might be able to utter only a single word.

One such person, a patient of Broca himself, was nicknamed "Tan" since that was the only sound that he could utter. Evidently, there was no problem with the vocal apparatus of these patients. On postmortem of these patients, Broca found that



**Fig. 1.8** Wilder Penfield explaining the maps he discovered using stimulation studies



**Fig. 1.9** A picture of the brain indicating Broca's and Wernicke's areas

they had a lesion in a part of the brain, usually located on the left hemisphere in right-handed individuals. This area, the so-called Broca's area was later found to be the key brain region that controls speech production (Fig. 1.9).

A related but contrary form of speech disorder was studied by Carl Wernicke, a German physician and psychiatrist. In the patients studied by Wernicke, speech is preserved but the sentences often do not make much sense. The level of impairment of speech may vary from a few incorrect or invalid words to profuse and unmixed jabberwocky. A patient suffering from this form of aphasia, named Wernicke's aphasia, may say, for example: "I answered the breakfast while I watched the telephone." The often heard expression of contempt toward another's intelligence—"He does not know what he is talking about"—perhaps aptly describes the speech of Wernicke's aphasics. Ironically, these individuals are painfully aware of their difficulty but there is little that they could about it. Patients with this form of aphasia also have difficulty understanding the speech of others. Thus while Broca's aphasia is an *expressive* aphasia, Wernicke's aphasia is described as a *receptive* aphasia. Postmortem studies revealed lesion in a part of the brain known as the inferior parietal area of the left hemisphere in right-handed individuals. This area was subsequently named after its discoverer as the Wernicke's area.

*Conduction aphasia* is a third form of aphasia related to the two forms of aphasia mentioned above. In conduction aphasia, patients have intact auditory comprehension and fluent speech. But their difficulty lies in speech repetition. Capacity for fluent speech suggests an intact Broca's area and a similar capacity for sentence comprehension indicates an intact Wernicke's area. Inability to repeat what is heard can arise when there is poor communication between Broca's area and Wernicke's area.

Indeed, conductive aphasia occurs due to damage of the nerve fibers that connect Broca's area and Wernicke's area.

The above examples of various forms of aphasia suggest a clear modularity in the distribution of brain functions. There is an area for speech production, another for speech comprehension and a connection between these two is necessary for speech repetition. Cognizing this modularity combined with interactivity, Wernicke presented a conceptual synthesis that solves in a single stroke the "local/global" debate that plagued functional neuroscience for many centuries. Wernicke proposed that though simple perceptual and motor functions are localized to single brain areas, more complex, higher level functions are possible due to interaction among many specific functional sites. For the first time, Wernicke drew our attention away from the "areas" to the "connections" and pointed out that the connections are important. Nearly a century after Wernicke, this notion of the importance of connections in brain function, inspired an entire movement known as "connectionism" and was expanded into a full-blown mathematical theory of neural networks. Out of such mathematical framework emerged concepts, jargon, a system of metaphor that can aptly describe brain function.

## Psychology

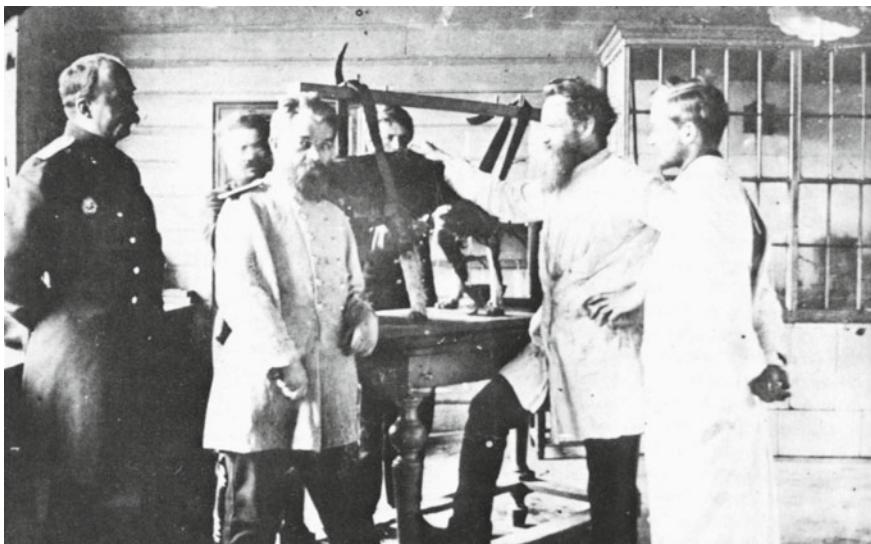
So far, we have seen how people learnt about various aspects of the brain: how neurons are shaped, how they are connected, how they converse among themselves by sprinkling chemicals on each other, how brain functions are distributed over various brain regions and so on. Obviously, there is a lot that could be said about each of these aspects and what was given above is only a very brief historical sketch. But even if a million details related to the above phenomena are given, the curious reader may not be really satisfied because most certainly a person interested in learning about the brain is not just interested in knowing about its intricate internal structures and manifold processes. Knowing about brain means, ultimately, to know how this mysterious organ creates and controls our thoughts, feelings, emotions, our experiences, or, in brief, our entire inner life. After reading a book on brain, one would like to know, for example, how we learn a new language, how we succeed (or fail) to memorize an immense amount of information by our desperate lucubrations on the night before a difficult exam, how we write poetry or appreciate music, how or why we dream, or how we live... and die? These definitely are samples of the most important questions about the brain that one would like to get answered.

These larger questions of our mental life are often the subject matter of psychology. In this field, elements of our inner life like thoughts, emotions, and even dreams are attributed a reality. But neuroscience takes a stricter stance, a stance that some might believe renders progress too slow and inefficient. That stance accepts only things one can "touch and feel," things that are concrete and measurable, quantifiable. But then is not this, fundamentally, the stance of all modern, Galilean science? Is not this simple yet formidable stance that has been the powerful driving force of all scientific

development over the past centuries? Thus, neuroscience seeks to explain every aspect of our mental life in terms of things that can be measured—neural activity, neurochemistry, concrete structural changes in the brain, and so on. Therefore, to explain purely in neural terms, why A has fallen in love with B, might be a tall order, even in the current state of neuroscience. One must start with some simple mental or behavioral phenomena to start and work one's way toward mind and emotions.

Since humans are already quite complicated, a group of psychologists who liked to have things concrete and measurable, decided to work with animals. They choose some very simple aspects of animal behavior, which, however, could possibly be related to their more sophisticated counterparts in humans. The simplest kind of behavior that can be studied is response to stimuli. The simplest kind of experiment would involve studying the cause-and-effect relation between a small number of stimuli and a small number of responses, where both stimuli and responses are measurable, quantifiable.

The famous, early class of experiments of this sort were the once performed by the Russian psychologist Ivan Pavlov. Like a lot of very impactful experiments, this one was an outcome of serendipity. Pavlov originally set out to study the physiology of digestion in dogs. He wanted to study all stages of digestion starting from the first one viz., salivation. It was common knowledge that hungry experimental dogs salivated when meat powder was presented to them. But Pavlov had the keen eye to observe that the dogs salivated even in presence of the lab technician who usually fed them. Based on this observation, Pavlov predicted that the dog will salivate in response to any stimulus that was consistently present when the dog was fed. Pavlov started to test this idea systematically over a series of experiments (Fig. 1.10).



**Fig. 1.10** Pavlov and his dog

In one such experiment, a bell was rung a short while before food was presented to the animal. Initially, the dog salivated in response to the food. But after repeated trials, the mere sound of the bell was sufficient to elicit a salivation response in the animal, which had, by then, learnt to associate the sound of the bell, with the presentation of food. This process of teaching the animal to respond in a predetermined fashion to a stimulus, which did not produce that response originally, is known as conditioning.<sup>1</sup> This is also one of the simplest forms of learning that can be studied at the level of behavior.

Experiments like those by Pavlov, and those who followed that tradition, inspired a whole school of thought in psychology known as behaviorism. Behaviorists took a slightly difficult, and impractical stance that all behavior can be—and ought to be—described in terms of observable quantities without invoking abstract and philosophical entities like mind. They denied existence to things like thoughts, feelings, insight, intelligence, and other elements of our subjective world and sought to explain the organism solely in terms of the responses of a “black box” brain to the stimuli from the environment. Behaviorism actually went even further. It did not even require data from the internal physiological processes of the body. It simply sought to build a science out of externally observable and measurable quantities—like decibels (of sound) and milliliters (of saliva). Notable among behaviorists are B. F. Skinner, Edward Thorndike, John Watson, and others.

Skinner practiced an extreme form of behaviorism known as “radical behaviorism.” His philosophy of research came to be known as Experimental Analysis of Behavior, wherein behavior is precisely measured and quantified and its evolution is studied. Skinner studied the role of reinforcement in shaping behavior. Reinforcements are rewarding inputs that modify behavior. Positive reinforcements are rewarding stimuli like food; negative reinforcements are punitive stimuli which the organism tries to avoid. An animal evolves responses that tend to increase the chances of obtaining positive reinforcements and reduce the occurrence of negative reinforcements. The process by which an animal acts/operates to maximize its reinforcement is known as operant conditioning.

Thorndike, like other behaviorists, confined himself to experimental methods and rejected subjective methods. He wanted to know if animals followed a gradual process of adjustment that is quantifiable, or used extraordinary faculties of “intelligence” and “insight.” He disliked the use of terms like “insight” that create an illusion of comprehension but explain nothing. Criticizing his contemporary literature on animal psychology he once said: “In the first place, most of the books do not give us a psychology, but rather a eulogy of animals. They have all been about animal intelligence, never about animal stupidity.” Out of the extensive experiments he performed with animals he deduced a few “laws” of learning:

- *The law of effect* stated that the likely recurrence of a response is generally governed by its consequence or effect generally in the form of reward or punishment.

---

<sup>1</sup>This form of learning, in which the involuntary response (salivation) of an animal to stimulus is studied, is known as classical conditioning. There is a very different class of conditioning known as instrumental conditioning, in which the animal produces a voluntary response.

- *The law of recency* stated that the most recent response is likely to govern the recurrence.
- *The law of exercise* stated that stimulus–response associations are strengthened through repetition.

Thus, Thorndike's work gave insight into how the associations or “connections” between stimuli and responses are governed by reinforcements, or are shaped by recent experience or practice. This study of stimulus–response connections becomes a concrete, well-defined problem at the heart of animal psychology. Tour de force attempts to extrapolate this approach to the rich, multihued world of human behavior ran into rough weather. But the trimming down of behavior, animal or human, to its bare quantifiable essentials has its advantages. Progress in the neuroscience of the later part of the twentieth century succeeded in finding a palpable, structural basis—a neural substrate—to the abstract connections that Thorndike and others dealt with.

The neural substrates to the abstract stimulus–response connections, interestingly, happen to be concrete connections between neurons—the synapses. Neurochemical modification of these synapses turns out to be the substrate for the evolution of stimulus–response behavior, often called *learning*. Thus, the notion that the synapse is a primary substratum for the great range of learning and memory phenomena of animal and human nervous systems, is now celebrated as one of the fundamental tenets of modern neuroscience.

## Summary

In this chapter, we rapidly traversed through some of the key historical ideas about the brain. We saw how certain misconceptions—for example, the idea of animal spirits controlling movement—stuck on for millennia until the modern times. We have also noted the tributaries of science that fed the great river of modern neuroscience. It must be conceded that history as it is presented in this chapter is far from being comprehensive, even in summary. The objective of this historical discussion is to glean certain key ideas of the brain, as they have emerged in history, and develop these ideas through the rest of the book.

What is sought in this sketchy presentation of history is to construct the picture of the brain as it emerged just before the current era of molecular neuroscience, biomedical technology, and computing. That essential picture is infinitely enriched and expanded by these recent developments, but the soul of that picture remains intact. Based on what we have seen so far, we make two important observations about the nature of the brain.

- (1) The brain is, first and foremost, a network. It is a network of neurons, or, a network of clusters of neurons. Each cluster, or a *module*, performs a specific, well-defined task, whereas, the performance of a more involved activity, like speaking, for example, requires coordinated action of many modules. Such a depiction of brain's processes is dubbed *parallel and distributed processing*.

(This synthesis, which is nearly identical to Wernicke's synthesis of brain function based on his studies of aphasias, resolves the longstanding *local vs global* conflict that raged in neuroscience for many centuries.)

- (2) Brain is a flexible and variable network. Since the network is constituted by "connections"—the synapses and the "wire"—structural and chemical modification of these connections gives the brain an immense variability.

Thus the brain presents the picture of a system that is a large, complex, and variable network. No other organ in the body fits this peculiar description. No wonder the brain occupies a position of pride in the comity of body's organs.

In the following chapter, we describe a different history of the brain: not a history of our ideas of the brain, but a history of the brain itself. It quickly traces the trajectory of the brain from its early beginnings in evolution, to its current position. Such a description of the brain in its primordial form might give an insight into its nature, which an intimidatingly detailed, textbook-like description of the human brain might not give.

## References

- Chisholm, H. (Ed.). (1911). Du Bois-Reymond, Emil. *Encyclopædia Britannica* (11th ed.). Cambridge: Cambridge University Press.
- Ehrlich, P. (1913). Address in pathology on chemotherapy: Scientific principles, methods and results. *Lancet*, 2, 445–451.
- Finger, S. (2000). *Minds behind the brain: A history of the pioneers and their discoveries*. Oxford: Oxford University Press.
- Graham, G. (2010). Behaviorism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2010 Edition). URL: <http://plato.stanford.edu/archives/fall2010/entries/behaviorism/>.
- Grmek, M. D. (1970–80). Claude Bernard. *Dictionary of scientific biography* (Vol. 2, pp. 24–34). New York: Charles Scribner's Sons.
- Gross, C. G. (1998). *Brain, vision, memory. Tales in the history of neuroscience*. Cambridge: MIT Press.
- Kandel, E. R. (1991). Brain and behavior. In E. R. Kandel, J. H. Schwartz, & T. M. Jessel (Eds.), *Principles of neural science* (3rd ed., pp. 5–17). Appleton: Lange.
- Lewis, J. (1981). *Something hidden: A biography of Wilder Penfield*. New York: Doubleday and Co.
- Patton, L. Hermann von Helmholtz. *Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/entries/hermann-helmholtz/>.
- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (G. V. Anrep, Trans.). London: Oxford University Press.
- Ramon y Cajal, S. (1906). The structure and connexions of neurons. In *Nobel Lectures: Physiology or Medicine, 1902-1921* (pp. 220–253), 1967. Amsterdam: Elsevier.
- Shepherd, G. M. (1991). *Foundations of the neuron doctrine*. New York: Oxford University Press.
- Van Leeuwenhoek, A. (1932). *Antony Van Leeuwenhoek and his "Little Animals"*. USA: Dover Publications Inc.
- Wernicke, C. (1908). The symptom complex of aphasia. In A. Church (Ed.), *Disease of the nervous system* (pp. 265–324). New York: Appleton.
- Young, R. M. (1970). *Mind, brain and adaptation in 19th century*. Oxford: Clarendon Press.

## Chapter 2

# Brain—Through the Aeons



*We are the product of 4.5 billion years of fortuitous, slow biological evolution. There is no reason to think that the evolutionary process has stopped. Man is a transitional animal. He is not the climax of creation.*

—Carl Sagan.

### The Anatomy of Intelligence

The last chapter was about the evolution of ideas of the brain. We have seen the war of two important views of brain, the “local vs. global” rivalry. We have noted Wernicke’s beautiful synthesis of our understanding of various aphasias: that simple functions (e.g., speech production) are localized in the brain, whereas more complex functions (e.g., speech in general) are performed by a coordinated action of several brain areas. Thus the brain, we were told, is ideally viewed as a parallel and distributed processing system, with a large number of events happening at the same time in various parts of the brain. With this basic understanding, it is perhaps time to take a peek into the wheels of the brain, and ask more specific questions: which parts of the brain process visual information? Which parts process emotions? And so on.

A standard move at this juncture would be, which is what a typical textbook on neuroscience would do, to launch a valiant exploration into the jungle of brain’s anatomy: the hemispheres, the lobes, the sulci and the gyri, the peduncles and fasciculi, and overwhelm the innocent reader with a mass of Greco-Latin jabberwocky. But this book is about *demystifying* brain. Therefore, taking the gullible reader on a daredevil journey through the complex, mind-numbing architecture of the brain is exactly what must *not* be done at this point. One must first get at the logic of that architecture. What is its central motif? What are its recurrent themes? What are the broad principles of its organization? Is there a pattern underlying that immense

tangle of wire? It is these insights that one must be armed with before we set out on a formal study of brain's anatomy.

One must remember that, paradoxically, even the most rigorous analysis of gross, material aspects of the brain, often has a sublime, immaterial objective: what is the structural basis of brain's intelligence? If mind, reason, and logic are what distinguish a human being from other species on the planet, intelligence is that prized quality that distinguishes often a successful individual from less successful ones. What features of a brain makes its owner intelligent? What deficiencies in the brain condemn a man to idiocy? If neural substrates of intelligence are understood, and if intelligence can be enhanced by direct neurochemical or surgical manipulation of the brain, it gives new teeth to the "IQ enhancement" racket. Thus, apart from the intrinsic philosophical and scientific import, there is an obvious market value to finding the answer to the question: what is the secret of brain's intelligence?

This was the question that drove Thomas Stoltz Harvey, a pathologist by training. It does not require a lot of intelligence to figure out that if you wish to study the secret of brain's intelligence, the simplest route would be to study an intelligent brain. That's exactly what Harvey did. He was fortunate enough to lay his hands on the brain of none less than Einstein, the mind (and the brain!) that is almost defining of twentieth-century science. After Einstein's death, Harvey performed autopsy, removed the brain, and preserved for a detailed study. Harvey hoped that a detailed study of the surface features of the brain, the precise arrangement of convolutions of the brain, the "hills and valleys", might give some clue.

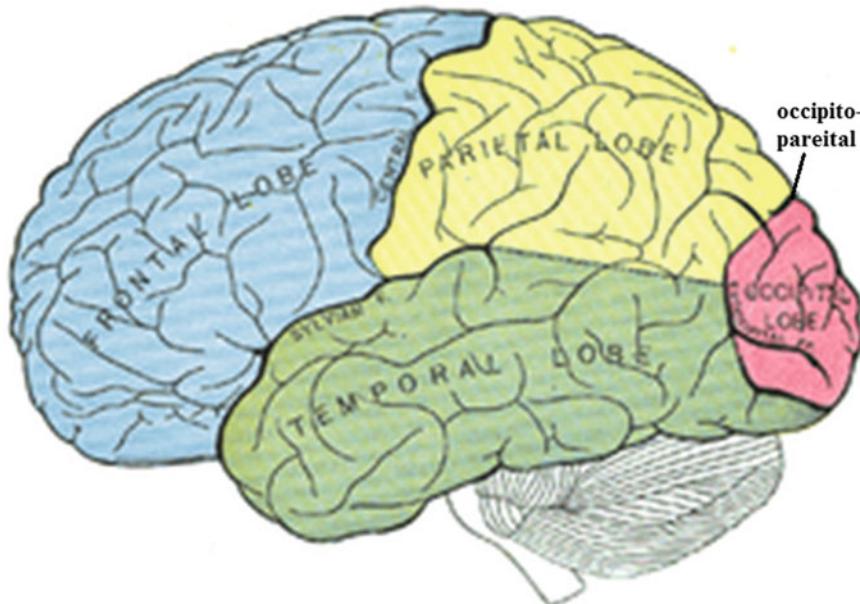
A live brain is soft and floppy like a jelly. In order to perform anatomical studies, or make sections for observing the internal components, the substance of the brain must be slightly hardened, or "fixed." Formalin is normally used as a fixation agent in brain studies. Harvey injected 10% formalin through the arteries of the brain, the network of blood vessels that deeply permeate the brain. A stiffer vascular network gives the brain a greater structural integrity. Harvey then suspended the brain in 10% formalin to obtain a more robust surface. He took pictures of the brain, thus prepared, from many angles. He then dissected the brain into about 240 blocks and encased these blocks in a substance called colloidin. Harvey then compared the anatomical features—at large scale and microscopic level—of Einstein's brain with that of an average one.

In order to appreciate the results of Harvey's studies, we must familiarize ourselves of preliminary topography of the brain. The brain and the spinal cord system are analogous to a bean sprout with its two split cotyledons comparable to brain's hemispheres, and the root comparable to the cord. Each of the hemispheres (the left and the right), has four large anatomically distinct regions known as the *lobes*. The lobe in the front, close to the forehead, is the frontal lobe; the one in posterior end, near the back of the head, is the occipital lobe; the large region between the frontal and the occipital is the parietal lobe; the region below (or inferior) to the frontal and the parietal is known as the temporal lobe. There are deep grooves that separate various lobes, named ornately as fissures, sulci (sulcus is Latin for furrow), and so forth. The boundary between frontal and parietal is the central sulcus. Sylvian fissure separates the parietal and temporal lobes. The boundary between parietal and

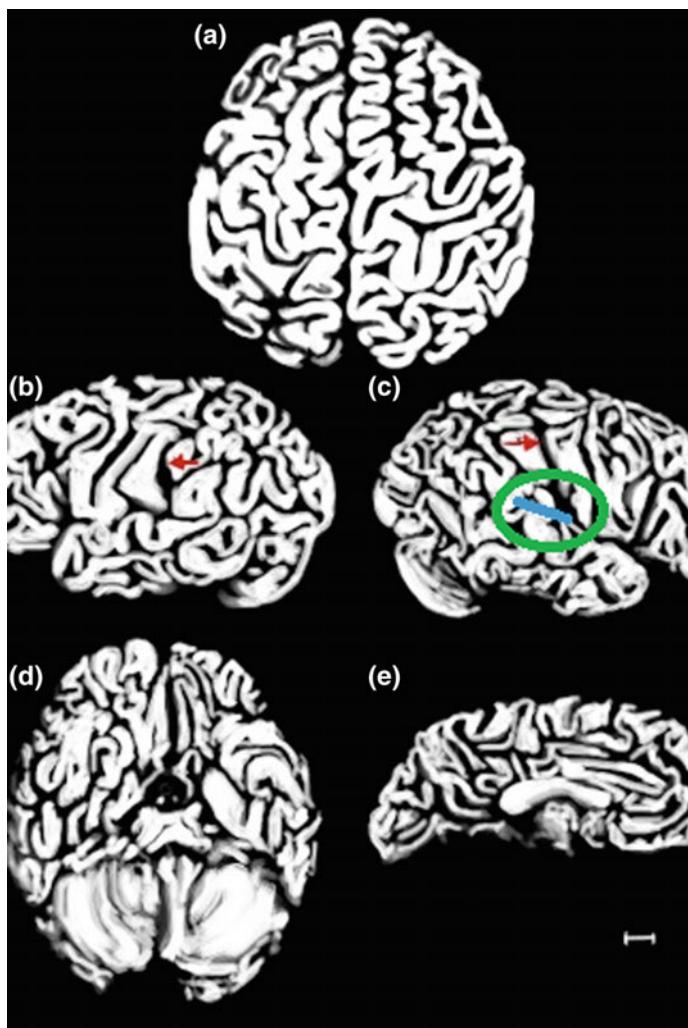
the occipital lobes has a more obviously pompous name—the parieto-occipital sulcus. Brain's surface is marked by distinct convolutions, the sulci ("valleys") that we just encountered, and the gyri (the "hills"). The arrangements of sulci and gyri in the human brain are nearly universal, and are therefore given standard names, though some individual variations do exist. The surface of the brain is a 2–5 mm thick layer of cells called the cortex (Fig. 2.1).

The differences that Harvey's studies found between Einstein's brain and an average brain are located around the inferior region of the parietal lobe and the Sylvian fissure. If you probe into the Sylvian fissure, you will arrive at a spot where three surfaces meet: one below, one above, and one right ahead. The one below is the part of the temporal lobe and the one right ahead is called the insula. Now the third surface, the one above, part of the parietal lobe, is known as the parietal operculum (operculum = Latin for "tiny lid"). This tiny stretch of the cortex is found to be missing in Einstein's brain. Another distinct feature is that the Sylvian fissure, which extends far beyond the central sulcus and separates the parietal and temporal lobes to quite some length, is much shorter in Einstein's brain. (The blue line inside the green oval in Fig. 2.2c indicates what would have been the Sylvian fissure extended into the parietal lobe in normal brains.)

In the 1980s, Marian C. Diamond of the University of California, Berkley, obtained some samples of Einstein's brain from Harvey and performed cellular level analysis on them. Diamond's team sliced the tissue samples into very thin slices each about



**Fig. 2.1** The brain and its lobes



**Fig. 2.2** Views of Einstein's brain

6  $\mu\text{m}$  (one-thousandth of a millimeter) thick and counted various types of cells found in them, with the help of a microscope. They found that Einstein's brain had slightly more of a certain type of brain cells called the glial cells in most parts. These glial (named after Latin for "glue") cells are usually greater in number than neurons in any brain. For a long time, they were thought to provide a structural support to neurons (like a "glue") or provide scavenging functions like clearing the debris of dying neuronal structures. But more recent findings about their involvement even in neuronal communications increase their status from being neuronal sidekicks to key, or even dominant, players in neural information processing mechanisms of the brain.

Einstein's brain had a particularly high concentration of glial cells in a specific part known as the association cortex. This part of the cortex, located in the inferior parietal lobe, combines information from the three adjacent sensory cortical areas (visual, somatosensory, and auditory) and extracts higher level abstract concepts from them. Einstein's brain had 73% more glial cells in the association cortex than in average brains.

Thus, though studies found clearly discernible, and statistically significant differences between Einstein's brain and the average run-of-the-mill variety, on the whole, the differences are unimpressive, almost cosmetic. A slightly shorter groove and a few extra cells in one little patch of brain surface seem to make all the difference between an idiot and a genius. This cannot be. Perhaps, we are missing something fundamental in our attempt to understand what makes a brain intelligent.

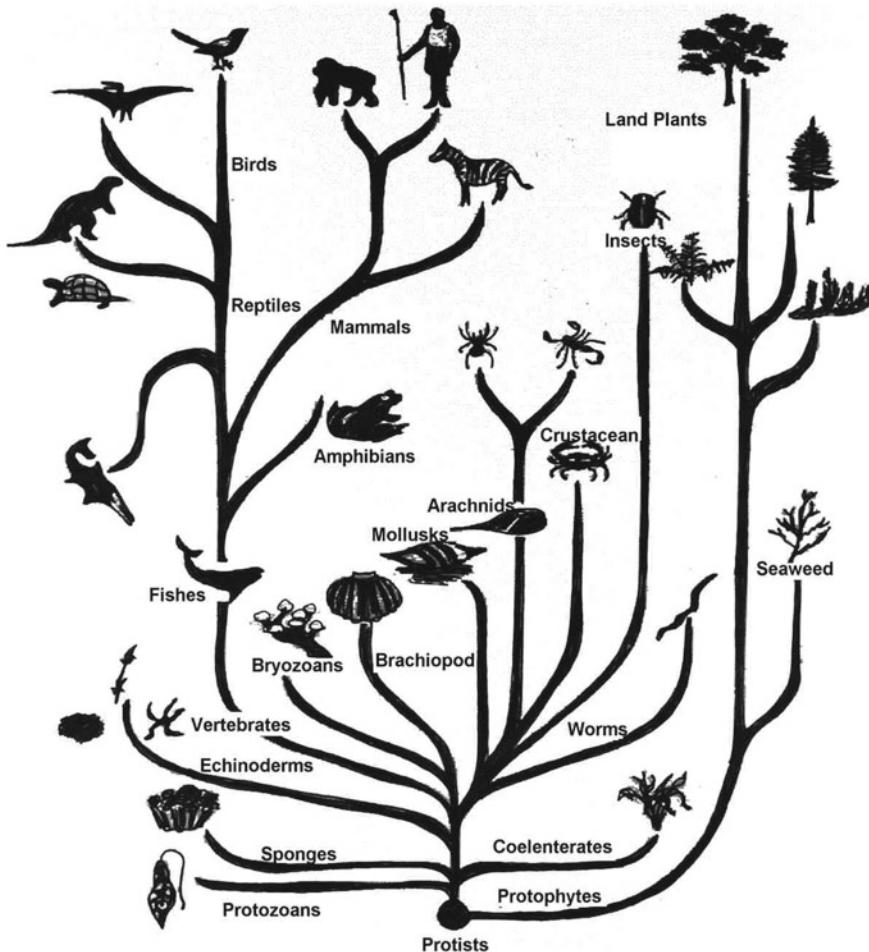
## The Evolution of the Nervous System

In the present chapter, we will focus on this passive or structural aspect of the nervous system, and attempt to draw some important lessons about the same. We will begin by looking at the structure of the nervous systems of very simple organisms, and see how the architecture of the nervous system grows more complex in larger organisms with a richer repertoire of behaviors. We will see that there is a logic and a pattern in that growth, a logic that is familiar to electrical engineers who design large and complex circuits and struggle to pack them in small spaces.

The creatures considered in the following section are: (1) hydra, (2) jellyfish, (3) earthworm, (4) octopus, (5) bird, (6) rat, (7) chimpanzee, and finally (8) the human. Figure 2.3 locates the above creatures in a simplified “tree of life.” Both hydra and jellyfish belongs to the phylum coelenterata shown in the bottom right part of the figure. The earthworm is located on the branch labeled “worms” slightly above coelenterate. Octopus belongs to the phylum mollusk shown close to the center of the figure. Birds are seen near the top-left corner. The rat, the chimpanzee, and the human are all mammals shown in one large branch in the top-center part of the figure. Let's consider the nervous systems of these creatures one by one.

### Hydra

It is a tiny (a few millimeters long) organism found in freshwater ponds and lakes under tropical conditions. In recognition of its incredible regenerative abilities, it is named after a creature from Greek mythology, the Lernean Hydra, a many-headed snake, whose heads can multiply when severed from its body. The tiny water creature hydra also shows some of the uncanny regenerative abilities of its mythological archetype. When small parts are cutoff a hydra, individual parts can regenerate into a whole animal. Even as early as 1740, Abraham Trembley, a Swiss naturalist, observed



**Fig. 2.3** A simplified tree of life. The eight species compared in the subsequent discussion can be located in the figure (see the text for explanation)

that a complete hydra can grow out of 1/1000th part of a hydra. People who studied the hydra's amazing ability to regrow and renew themselves wondered if these creatures are actually immortal. The question of verifying the immortality of hydra is a bit tricky, considering that those who would perform such studies would themselves be necessarily mortal. But studies that traced the development of a population of hydra over a span of 4 years, noticed no signs of senescence. For all that we know, the hydra might be truly immortal.

But the aspect of hydra which we are particularly interested is its nervous system. Hydra has simple nervous systems that govern their curious locomotive behaviors. Figure 2.4 shows the body of a hydra with its tree-like structure, and its "branches"

**Fig. 2.4** Hydra

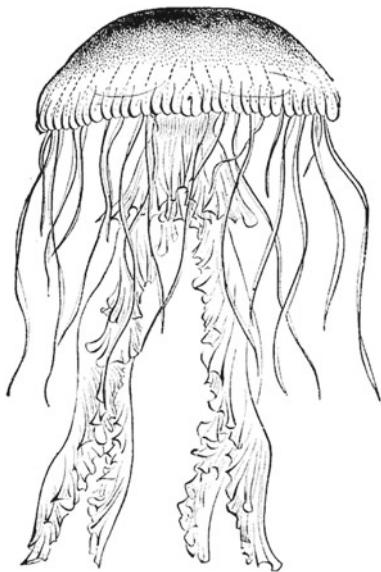
known as tentacles. These tentacles, the hydra's arms, double up as feet too. At the base of the tentacles, or top of the stalk, the hydra's mouth is located; through this orifice, hydra ingests food and also expels waste matter. When a hydra is attacked, alarmed, or plainly upset, its tentacles can retract into small buds, or its entire body can roll itself up into a gelatinous ball. A hydra is usually sedentary but when it has to hunt, it moves around by "somersaulting." It bends over to a side, adheres to the substrate by its tentacles and mouth, lifts its foot off, and goes topsy-turvy. By making such movements in a cyclical fashion, it slowly inches along a surface, moving by as much as several inches in a day.

Hydra's movements can be generated and coordinated by a nervous system consisting of a diffuse network of neurons, known as the *nerve net*, distributed all over its body. Its nervous system is not an aggregated mass of neurons, like the brain or spinal cord that we possess. Hydra has specialized cells that detect touch and also the presence of certain chemical stimuli. These signals spread over its nerve net, causing appropriate convolutions throughout its body and tentacles.

A similar nerve net governs the life of another creature that lives in a very different milieu, exhibit very different behaviors.

## Jellyfish

Jellyfish belong to a family of organisms known as the plankton, which inhabit upper layers of oceans, seas, and freshwater bodies (Fig. 2.5). The word plankton comes from Greek planktos (a root from which the word planets, the "wanderers", comes from), which means "wandering." Plankton typically drifts with water cur-

**Fig. 2.5** Jellyfish

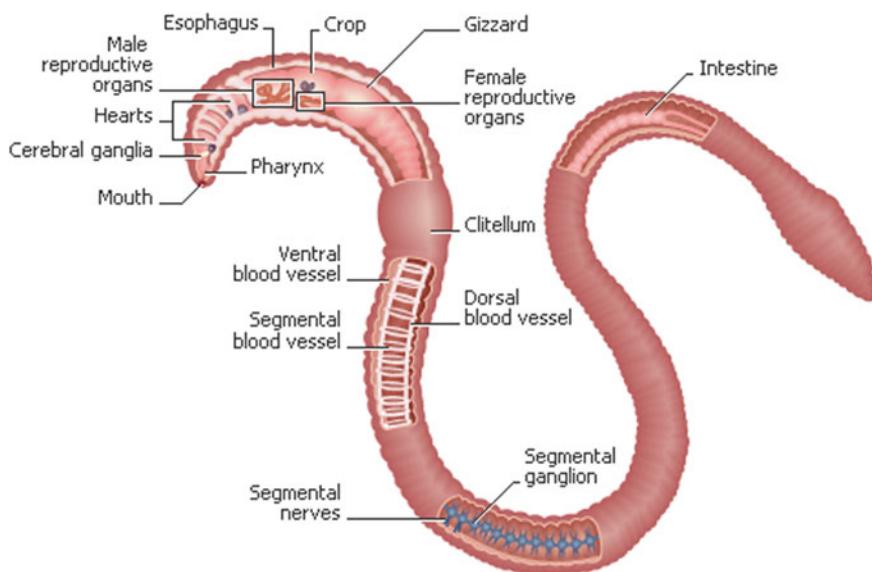
rents, though some are endowed with a limited ability to swim and move around. Jellyfish are some of the largest forms of plankton. Like hydra, jellyfish too has tentacles which are used to catch and paralyze food, and carry to their large stomachs. The jellyfish uses its stomach for locomotion, by pumping water with its stomach. However, this procedure can mostly carry it in vertical direction, while for horizontal transport, it simply depends on the currents. Some jellyfish, like the *Aurelia*, for example, have specialized structures called *rhopalia*. These structures can sense light, chemical stimuli, and touch. They also give the jellyfish a sense of balance, like the semicircular canals<sup>1</sup> in our inner ears. All this sensory-motor activity of the jellyfish is driven by a simple nervous system, a nerve net similar to that of a hydra. Surely, a great evolutionary distance separates the diffuse nerve net of the jellyfish, from the brain and spinal cord in humans. Between these two extremes, there are intermediate stages. Let us consider a nervous system that is a step above the diffuse nerve net in complexity.

---

<sup>1</sup>These are fluid-filled rings located in our inner ears with a role in maintaining our balance. When our heads spin suddenly, as they might when we are about to lose our balance, the fluid in these canals flows past an array of sensors inducing electrical signals. These signals are used by the brain to initiate corrective measures and restore balance.

## Earthworm

This creature of the soil seems to have a nervous system with a structure that is one step above that of the diffuse nerve nets we have encountered in the previous two examples (Fig. 2.6). Neurons in the earthworm's nervous system are not distributed loosely but clumped into structures called the *ganglia*. These ganglia form a chain that extends along the linear, snake-like body of the earthworm. At the "front" end of this chain, there exists a special mass of neurons—the cerebral ganglia (Fig. 2.6), which come closest to what we think of as a brain. The cerebral ganglia are connected to the first ganglion, known as the ventral ganglion. Although the earthworm's nervous system is slightly more structured than a diffuse nerve net, it does not have the wherewithal to provide a coherent, unified control of the body like more advanced nervous systems. In an earthworm's nervous system, meaningfully classified as a segmented nervous system, different ganglia control only a local portion of the body of the earthworm. The brain itself has a key role in the control of the earthworm's writhing movements. If the brain is severed from the rest of the nervous system, the organism will exhibit uninhibited movement. Similarly, severance of the ventral ganglion will stop the functions of eating and digging. Other ganglia also respond to sensory information from the local regions of the body and control only the local muscles.



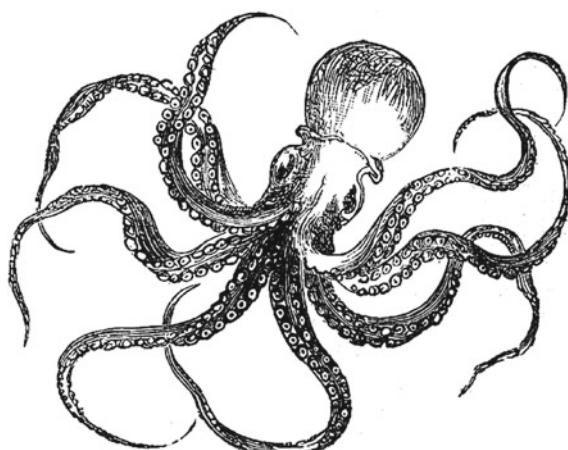
**Fig. 2.6** Illustration of an earthworm showing segmental ganglia

## Octopus

The previous three cases considered are examples of invertebrates, which have simpler nervous systems, and therefore more limited repertoire of behaviors, compared to their evolutionary successors—the vertebrates. But there is one creature, an invertebrate, (since it does not possess a backbone, but more precisely, it is categorized as a mollusk) which may have to be placed on the highest rungs of invertebrate ladder—the octopus (Fig. 2.7). This creature exhibits an impressive range of complex behaviors bordering on what may be described as intelligence. The octopus inhabits the oceans, with particularly high concentrations found in the coral reefs. It has eight arms or tentacles which it uses for a variety of purposes including swimming, moving on hard surfaces, bringing food to its mouth, or attacking the prey. Although lacking a bony skeleton from which the arms of vertebrates get their special strength, the octopus' tentacles are remarkably strong enabling them to wrestle with sharks or breakthrough plexiglass. The tentacles are lined with an array of suction cups, with which the animal can attach itself to surfaces. Octopus's eyes are similar to ours in structure—with iris, pupil, lens, and retina—and have inspired manufacturers of cameras. Study of the lens in octopus' eye led to improved designs of the camera lens. Traditional cameras were using homogeneous lenses, curved at the edges. Due to this curvature, the images formed are often blurred at the edges. But analysis of octopus' lens led the manufacturers to make lenses with several layers of varying densities, greatly improving the image quality.

Apart from these sophisticated bodily bells and whistles, the octopus is remarkably intelligent for an invertebrate. Like a kindergarten child, it can distinguish simple patterns and shapes. It is found to be capable of “playful” behavior, something like repeatedly releasing bottles or other toys in a circulating stream and catching them again. They were also seen to be able to open a container with screw caps. In one study conducted in Naples, Italy, an octopus learnt to choose a red ball over a white ball by

**Fig. 2.7** Octopus



observing another octopus. The researchers were shocked to see such “observational learning,” the ability to learn by watching another organism, in an octopus. Because such capability is often seen in animal much higher up on the evolutionary ladder, like, for instance, rats. A recent case of octopus intelligence was the performance of Paul, an octopus used to predict match results in World Cup Soccer (2010). This gifted octopus was able to predict the results of every match that Germany had played, and also the final winner of the cup. This ability is not just “intelligent” but borders on the “psychic” considering that the odds of the predictions coming true are 1:3000. While the “psychic” side of an octopus’ personality is rather difficult to account for, the other intelligent activities of this wonderful creature fall well under the scope of neurobiology. Like the other invertebrates we visited in this chapter, the octopus too has a nervous system that consists of a network of ganglia. One of these located in the “head” is called the brain, which consists of only a third of the neurons in its nervous system. The tentacles are also controlled by separate ganglia and therefore have nearly autonomous control. Let us try to put the brain of an octopus in an evolutionary perspective.

The octopus is a cephalopod mollusk, a subclass of invertebrates. A mollusk is a kind (a phylum) of invertebrates that lives mostly in seas and freshwater habitats, though there are some that live on the land too. The nervous systems of this class of organisms consist of two chains of ganglia running along the length of the body, like railway tracks. In the cephalopods alone (of which the octopus is an example), among the mollusks, evolution has created a brain. The forward most pairs of ganglia are expanded and brought together to create tightly packed mass of neurons, called the brain, which is located behind the ears, encircling the esophagus. Thus, the nervous system of a cephalopod mollusk like the octopus, is a nervous system in transition, from the chains of ganglia of invertebrates to one with a brain and a spinal cord in the vertebrates. Let us now consider some patterns of development of vertebrate nervous system.

The vertebrates are not a single monolithic group, but a massive family with a large number of branches and subbranches. For example, at a high level, the vertebrates are classified into jawed and jawless vertebrates; the jawed are further divided into bony vertebrates and cartilaginous fishes; the bony vertebrates are again classified into lobe-finned vertebrates and ray-finned fishes, and so on. If we step down the line of lobe-finned vertebrates and step down a few branches, we arrive at the amniotes (four-footed creatures with backbone, which emerge from an egg that can survive on land). As we continue down the line of amniotes, we successively encounter mammals, placentals, primates, and humans. It would indeed be a brash and sweeping statement to say that the nervous systems of the great vertebrate branch of life are endowed with a brain and spinal cord, and other key structures like cerebellum. But the precise evolutionary changes in the nervous systems as one ascends the ladder of evolution, rung by rung, are the preoccupation of an evolutionary biologist. Our present purpose is only to see certain broad trends in the development of the gross structure of the nervous system, and perceive a lucid logic that is grounded in the physical principles that govern such development. Thus, we will satisfy ourselves

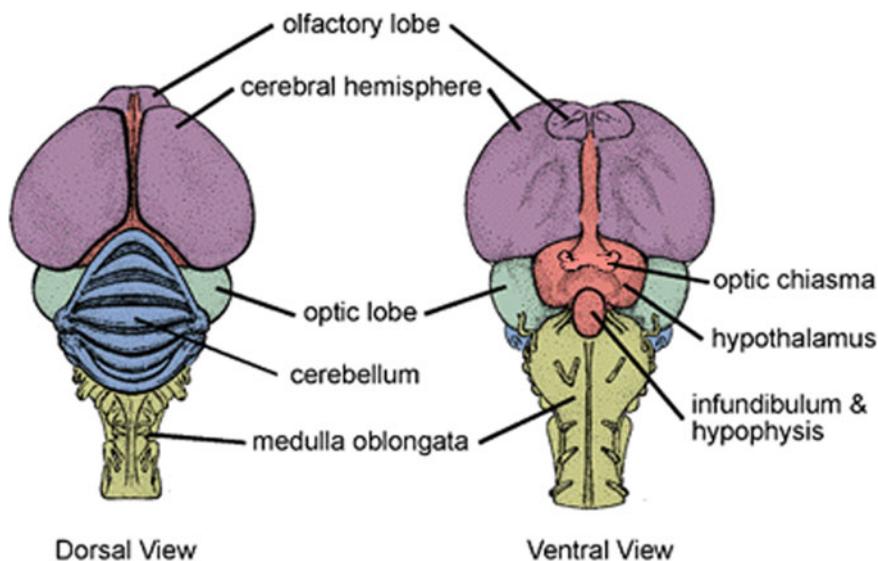
with a brief description of nervous systems of a few vertebrates and the capabilities that those nervous systems bestow on their owners.

## Songbirds

Although considerably low in vertebrate hierarchy, birds already exhibit a remarkable type of intelligent skill—the birdsong. A fully developed, intelligible, intelligent speech and language may be the exclusive privilege of humans. But in the animal world, the song of the songbird is perhaps something that comes closest to human speech. Though the birdsong formed part of literature and poetry for millennia, a systematic scientific study of birdsong is only a couple of centuries old. Even as early as 1773, it was clearly demonstrated that young birds learn their song from older birds—the tutor birds. The earliest record of a birdsong was done by Ludwig Koch in 1889, which he did when he was eight. The song that he recorded was of a bird known as the Indian Shama, a member of the thrush family. In his *Descent of Man*, Charles Darwin discussed some of the nonvocal sounds of birds—like the drumming of woodpeckers, for example—and described them as instrumental music. Apart from these voiceless sounds, birds make unique, distinctive sounds with their syrinx, an avian equivalent of the human larynx. Just as the colorful plume and feather, a bird that is capable of singing—the songbird—can also be identified by its song. Songs are distinguished from calls, which are usually shorter. While birdsongs are often associated with courtship and mating, calls are used for raising alarm or keeping the members of the flock in contact.

Sometimes, the songs can be so unique to the individual bird, that birds identify each other through their song. Birds that nest in colonies often identify their chicks with the help of their calls. Some birds tend to respond to each other with reciprocating calls, a habit called *duetting*. In some cases of duetting, the calls are so perfectly timed, that they seem to be a single call. Some birds are excellent at mimicking vocal sounds, usually those of the same species though sometimes they can imitate the songs of other species also. This mimicking quality of birds often comes as a source of amusement to rural children, who enjoy themselves by imitating the call of a cuckoo and provoking the unwitting bird into a quaint homo-avian *jugalbandi*.

Just as an aficionado learns music from a virtuoso by years of practice, the songbird learns its song from a tutor bird over an entire season. A lot of our understanding of birdsong came from studies of a songbird called the zebra finch. Young chicks of zebra finch learn to sing a crude version of the song by about 20 days from the hatch. Song learning reaches a plateau by about 35 days. The bird now enters a “plastic” stage, which extends over a few months, during which the bird continues to perfect and fine-tune the song. Subsequent to the plastic stage, the song becomes robust and mature in the so-called “crystallized” stage. In some birds, like the zebra finch, learning is confined to the first year. But others like canaries are lifelong learners; they continue to learn new songs even as sexually mature adults.



**Fig. 2.8** Key structures in a typical avian brain. [http://www.uoguelph.ca/zoo logy/devobio/210labs/ecto3.html](http://www.uoguelph.ca/zoology/devobio/210labs/ecto3.html)

Singing is perhaps not the smartest thing that the birds do. Studies found that pigeons are capable of learning and memorizing 725 different visual patterns and can differentiate patterns into “natural” versus “human-made.” Some bird species are found to be capable of using tools, a skill that has been thought for long to be an exclusive right of higher mammals. The wedge-tailed eagle has been observed to break eggs using small rocks. The green heron has been seen to throw tiny crumbs of bread as bait to catch fish, showing a cunning that has been believed to be a uniquely human trait. Anecdotal evidence shows that birds can even count: crows up to three, parrots up to six, and cormorants all the way up to eight.

Considering such phenomenal cognitive abilities of birds, it would be no more an insult to call someone a “bird brain.” These uncanny abilities were made possible because the nervous system of the bird has come a long way from that of the invertebrates, with their unremarkable chains of ganglia. Some of the key structures seen in the human brain—like the cerebrum, cerebellum, basal ganglia, medulla, and spinal cord—are already seen in their primitive forms in the bird brain (Fig. 2.8). Evolutionary biologists have been faced with the conundrum of explaining how birds, which are considered “lesser” than mammals, are endowed with such cerebral abilities. Mammalian brains are particularly gifted by the presence of neocortex, a convoluted sheet of cells a couple of millimeters thick. The neocortex is thought to be the secret of the superior abilities of mammalian brains in general, and that consummation of neural evolution, the human brain, in particular. Recent studies have noted that the bulging portion in the front of bird’s brain, the forebrain, or more formally known as the pallium, has remarkable similarities, in terms of circuitry, cell types, and pres-

ence of neurochemicals, to the corresponding elements of the neocortex in mammals. Therefore, perhaps the pallium of bird brain is an avian version of the mammalian neocortex.

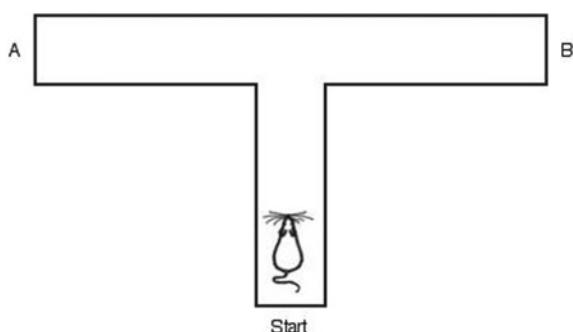
Now, let us consider the nervous system of a true mammal, and learn about the superior capabilities it affords.

## Rat Intelligence

In literature and popular media, rats are often associated with filth and pestilence, and described unfairly as lowly and despicable creatures. As mammals, though they are endowed with neural machinery which gives them significant cognitive capabilities, popular accounts of rats often underestimate their intelligence. One capability that rats are particularly known for is their skill in dealing with space. Rats' legendary mastery of space, as manifest in their ability to find their way around in mazes, makes them important subjects in neuroscience research. Maze learning by rats has been studied for nearly a century. When humans solve maze puzzles, they do so by looking down on the maze from "above." But solving a maze, by being a part of it, as a rat does it, is not trivial. After repeated exploration of the maze, the rats are known to build an internal model of their spatial surroundings in their brains, and learn to represent the route to the goal in terms of that internal model. Study of this ability of rats to build "route maps" of the spatial world provided valuable insights into how humans perform similar functions on a more elaborate scale.

In a simple instance of maze learning, a rat is placed in a T-maze (Fig. 2.9), with the rat always starting from the tip of the "stem", and a food reward always placed at the end of the arm pointing westward. Once the rat learnt the location of the food, the maze is rotated by 180°, in order to see if the rat was making stereotyped body turns ("turn left for food") or is it choosing a general direction ("turn westwards for food"); the rat was found to choose the latter strategy, proving that the animal is capable of representing the locations in the maze, in terms of landmarks located in the larger world outside the maze.

**Fig. 2.9** A rat in a T-maze



Rats too, like the birds, for example, are credited with some level of numeracy, or familiarity with numbers. For example, rats could discriminate between two, three, or four auditory signals presented in a sequence. They could be taught to take a fixed number of food pellets, say four, not more, not less, from a plate of pellets, by delivering punishments for picking a wrong number. Rats were trained to discriminate the number of times their whiskers were stroked. They could also exhibit an understanding of ordinality (“first, second, third”, etc.). In one study, rats were trained to always enter one of the six tunnels, irrespective of the absolute location of the tunnels, their appearance, odor, and other attributes.

Beyond performing successfully in these standard tests of intelligence administered to several other species, rats display other unconventional forms of intelligence. It was observed that a type of rat, known as the wood rat, hoards fresh leaves of a variety of plants. The researchers were surprised to find that the rats themselves do not eat these leaves. A careful investigation into the possible purpose of these leaves to the rats revealed that the leaves picked by the rats help to reduce the hatch rate of flea eggs in rats' sleeping nests. In other words, the rats have learnt the use of pesticides to keep bugs off their beds!

## Chimpanzee Intelligence

In any discussion of the evolution of intelligence, chimpanzees occupy an important place since they share the same family—the hominids—to which humans also belong. Another feature that marks the chimps as candidate creatures with a possible gift of intelligence, is that human beings and chimpanzees share about 99% common DNA, the molecule that constitutes our genetic material.

Use of tools as an extension to the natural tools like extremities, for example, is considered a sign of advanced intelligence in the animal world. We have seen earlier how even birds are capable of a certain elementary form of tool use. Research

**Fig. 2.10** A chimp using a twig as a tool to fish out termites



shows that chimps truly excel in this respect since they were found to be capable of using tools in over 350 different ways. One of the first instances of the tool used by chimpanzees was observed by Jane Goodall, an international expert on chimpanzee studies who dedicated her life to the study of these very special creatures. In one of her early visits to Africa, Jane Goodall began to study these creatures with the hope of finding something new, something her predecessors could not observe. Then, she observed a chimp using a thin stick as a line of some sort to fish out termites from a termite nest (Fig. 2.10). Termite hunting forms an important part of the life of a chimpanzee. Chimps were found to use thicker sticks to dig and make inroads into a nest, while slender sticks were used for “fishing” termites out. Thus, chimpanzees were described as being capable of using, not just tools, but a whole “tool kit.” Subsequent studies of chimpanzee behavior discovered a large number of other forms of tool use. For example, chimps could sharpen sticks and use them as spears in the fight. They could use leaves as cups and drink water from a pool. Or they could use a long slender branch to enjoy a game of tug-of-war with their mates.

Like in the case of birds and rats, chimpanzees were also found to possess a remarkable ability to work with numbers. In a famous study performed at Kyoto University’s Primate Research Institute, chimps were trained to recognize numbers from 1 to 9. These numbers were displayed at random positions on a computer screen. The animals were trained to touch these numbers in a sequence 1, 2, 3..., etc. Every time they successfully completed the task, they were rewarded by a nut or two. Once they reached a certain level of mastery on this task, the task was made even harder. This time, the animal was allowed to look at the display of numbers for a fraction of a second before the numbers were covered by white squares. Chimpanzees were expected to hold the numbers and their positions in their heads, and touch the positions in the correct sequences. Some chimpanzees were able to perform even this second task with astounding accuracy and speed. But what is even more startling and somewhat disturbing is that human subjects, graduate students to be precise, failed miserably on the second task which requires the subject to retain a photographic memory of the displayed digits.

Language is one of the highest faculties that a creature aspiring to be considered intelligent may possess. Most animals are capable of producing vocalizations of many sorts—shrieks and screeches, snarls, and growls. These can even carry a meaning like the distress call of an animal in trouble or the growl of an animal marking its territory. But these are not rich enough to be described as communication, far less a language. For a language or a system of communication implies a shared set of sounds and/or gestures which an animal deliberately uses to convey something to another animal and receives responses in the same code. By this definition, chimps certainly pass the test of a species possessing the ability to communicate.

Many important observations regarding the ability of chimps to communicate have come from a troop of wild chimps at the Gombe Stream Reserve on the shores of Lake Tanganyika. Chimps could direct the attention of a fellow chimp at a distant object by stretching the hand and pointing. They would raise their hands, like kindergarten children, drawing the attention of visitors toward themselves, and begging for food. They were also found to use hand gestures and symbols to communicate with each

other. These initial observations in the ‘60s and the ‘70s triggered a lot of research in primate communication. Efforts were made to deliberately train apes in using sign language.

Beatrix and Allen Gardner at the University of Nevada in Reno trained a chimpanzee called Washoe in usage of the American Sign Language (ASL). In the initial stages, Washoe learned 132 different words. Subsequently, 4 other chimps were also trained to sign. These 5 chimps, now began to move together like a family, and started to sign to each other—apart from humans—to communicate. Furthermore, Washoe taught the sign language to its adopted son, a baby chimpanzee, without any human intervention.

These creatures which were trained on ASL, were able to use signs, not just to denote individual objects but to whole families. For example, they would use the sign “dog” to denote all types of dogs. Most remarkably, they were able to combine familiar signs to denote new concepts, as, for example, a “drink fruit” which denotes a watermelon.

## Earmarks of a Smart Brain

We have sampled, albeit sparsely, the ladder of the species and considered the capabilities afforded to them by the nervous systems that they possess. Hand in hand with the evolution of form, and evolution of nervous systems, there is an obvious evolution of intelligence. Smarter animals do seem to have larger brains than less privileged ones. A pigeon that can be trained to classify simple visual patterns is definitely smarter than an earthworm with its primitive reflexes. A chimpanzee that can learn a whole sign language from a human tutor and teach it to its offspring, is certainly more intelligent than a crow that can count up to 3. Smarter animals indeed seem to have larger and more complex brains. But if we wish to go beyond these generalities and seek to make a more precise statement we quickly run into rough weather. Because in answering the question we need to answer two smaller questions: “what is a complex brain?” and “what is intelligence?”

It is vexing enough to define “complexity” of brain, and “intelligence” of an organism independently, not to speak of their interrelationship. How do we define complexity? The number of neurons, or the number of ganglia, or the number and size of different substructures of a mammalian brain? Similarly, one of the most perplexing aspects of comparing intelligent behavior across species is the absence of a common yardstick. The intelligence of an organism manifests itself in a milieu unique to that organism. Hence, it is not possible to test an animal’s intelligence independent of its milieu.

But let us stubbornly persist with our question and see if we can at least make some general observations about the distinguishing features of a smart brain. Let us begin with brain size or weight. Are larger brains smarter? It does not take much to negate this question. Both elephants and whales, of different varieties, have brains

**Table 2.1** Brain weight/body weight ratios for various species

| Species     | E/S ratio | Species      | E/S ratio |
|-------------|-----------|--------------|-----------|
| Small birds | 1/12      | Lion         | 1/550     |
| Human       | 1/40      | Elephant     | 1/560     |
| Mouse       | 1/40      | Horse        | 1/600     |
| Cat         | 1/100     | Shark        | 1/2496    |
| Dog         | 1/125     | Hippopotamus | 1/2789    |
| Frog        | 1/172     |              |           |

Kuhlenbeck (1973)

larger than ours. Therefore, sheer size may not hold the secret. The other possibility is to consider brain size (or weight) relative to body size (or weight).

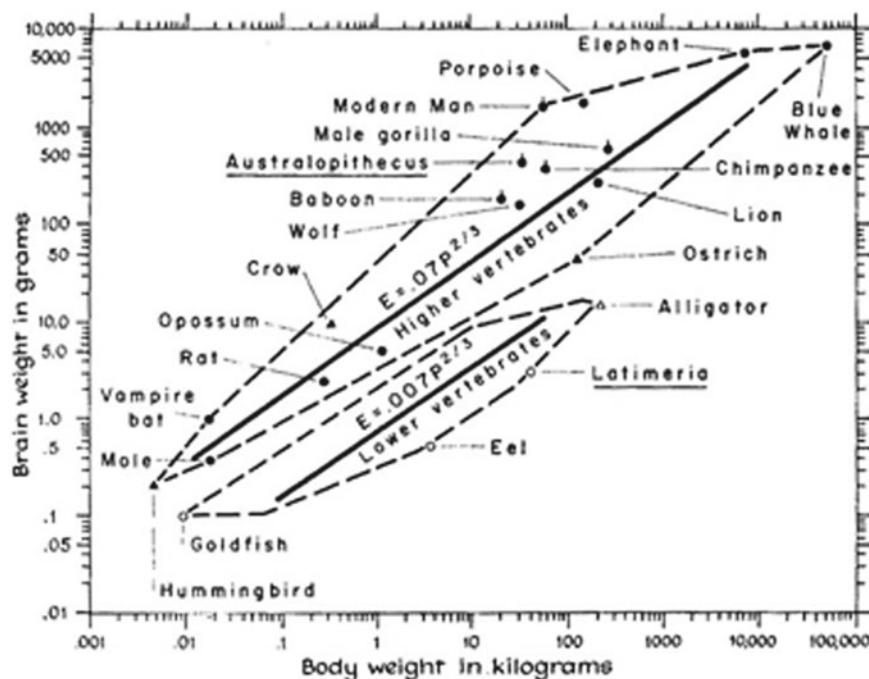
Therefore, it appears to be more logical to consider relative weight of the brain with respect to the body than absolute brain weight. Table 2.1 gives a list of relative brain weights of various species. This quantity, known as Cuvier fraction, is denoted by  $E/S$  where  $E$  (“encephalon” or brain) stands for brain weight and  $S$  (“soma” or body) for body. Note that, birds have the highest  $E/S$  values with humans following right after. It might be a bit disconcerting to note that humans and mice have the same  $E/S$  ratio. Lions and elephants have similar  $E/S$  ratios, as is the case with sharks and hippopotamus. Therefore,  $E/S$  does not seem to match up with our intuitive expectation of what a “smart” brain must be. There is a need to search for more meaningful anatomical parameters that might correlate with intelligence.

Neuroanatomists have indeed found a different parameter called, Encephalization Quotient (EQ), which emerges out of a discovery of a more logical, consistent  $E$  versus  $S$  relationship that holds good for a large class of species at the same time. Figure 2.11 above shows a plot of  $E$  versus  $S$  in a log–log plot. Two separate straight line plots are shown: one for higher vertebrates including mammals (the upper graph) and other for lower vertebrates (the lower one). Two polygons can be seen encompassing the  $E$  versus  $S$  points of higher vertebrates and those of lower vertebrates separately. These are known as “maximum polygons” which are the smallest/tightest polygons that encompass a given set of points on a plane. Now, a straight line relationship in a log–log plot actually means that the two quantities are exponentially related. Figure 2.11 shows two different exponential relations as follows:

$$E = CS^{2/3}, \text{ where } C = 0.07 \quad (\text{for higher vertebrates}),$$

$$E = CS^{2/3}, \text{ where } C = 0.007 \quad (\text{for lower vertebrates}).$$

In the above equations,  $S$  is measured in kilograms and  $E$  in grams. The power  $2/3$  was paid special attention by neuroanatomists. If we assume all animal bodies to have exactly the same geometry (which is obviously not true) and a fixed body density (not true too), then  $S^{2/3}$  is vaguely related to surface area of the body. But since there



**Fig. 2.11** Brain weight ( $E$ ) versus body weight ( $S$ ) in log–log scale

are so many assumptions underlying this interpretation, it is doubtful whether this interpretation can be taken seriously. Another reason for not taking the power value of  $r = 2/3$  seriously is that the exact value obtained depends on the choice of points on the  $E$  versus  $S$  plot. For example, Kuhlenbeck (1973) suggests a value of  $r = 0.56$  for mammals. Presently, we will choose  $r = 0.66 = 2/3$  and consider the two plots of Fig. 2.10.

Thus, a strategy for comparing brains of different species, or groups of species, is as follows. Choose the same power value,  $r$ , for different species and calculate the  $C$  value that gives the best fit between  $E$  and  $S$  of the form ( $E = CS^{2/3}$ ). The  $C$ , thus obtained may be regarded as a more sophisticated form of the simple  $E/S$  ratio we considered earlier. Now, this quantity ( $C_{\text{species}}$ ) is compared with the corresponding  $C$  value obtained for mammals ( $C_{\text{mammal}}$ ). The ratio of the two is known as Encephalization Quotient, and EQ is given by

$$\text{EQ} = (C_{\text{species}}) / (C_{\text{mammal}}).$$

Now, if a given species has EQ of 2, it means that its  $C$  value is twice that of an average mammal, or if it's EQ is 0.5, it means that its  $C$  value is half of that of an average mammal. Table 2.2 shows the EQ values of a range of species (Macphail 1982).

**Table 2.2** Encephalization Quotients of various species (Macphail 1982)

| Species       | EQ   | Species | EQ   |
|---------------|------|---------|------|
| Man           | 7.44 | Cat     | 1.00 |
| Dolphin       | 5.31 | Horse   | 0.86 |
| Chimpanzee    | 2.49 | Sheep   | 0.81 |
| Rhesus monkey | 2.09 | Mouse   | 0.5  |
| Elephant      | 1.87 | Rat     | 0.4  |
| Whale         | 1.76 | Rabbit  | 0.4  |
| Dog           | 1.17 |         |      |

The new parameter, EQ, seems intuitively far more satisfactory, than the simple ratio of E/S. Species that are known for their intelligent behavior—dolphin, chimpanzee, monkey, including man—are associated with the largest values of EQ. Thus, we were able to consider some data related to gross brain weight of various species, extract a parameter related to brain weight relative to body weight, and show that the parameter approximately correlates with the position of the species on the ladder of evolution. But in spite of all this exposure to evolutionary anatomical data, we must confess that we still do not have an *insight* into what exactly makes a brain smart. What anatomical earmarks make a brain produce intelligent behavior?

It is not difficult to see that in order to answer the last question, it is not sufficient to consider gross anatomical facts about the brain. One must look at the internal structure of the brain, the logic of organization of the brain, and try to seek answers at that level. Because in our discussion of nervous systems of different organisms from hydra to chimpanzee, we noted that, in addition to the simple trend of growing brain weight, there is also an evolution in brain's organization. The trend may be described as diffuse nerve net (e.g., hydra, jellyfish) → chains of ganglia (e.g., earthworm, octopus) → brain and spinal cord (mammals like us). Does this organization have anything to do with substrates of intelligence? For a fixed body weight and brain weight, can we say that a nervous system with a compact brain and spinal cord is smarter than one with a diffuse nerve net? How do we even begin to answer such questions? To start with, we learn about an important evolutionary principle that governs the logic of brain's organization.

## The Logic of Brain's Organization

In the previous section, we made a whirlwind tour of the evolution of the brain by quickly visiting a few interesting milestones on the way. Starting from a primitive organism like the hydra, and climbing all the way to the chimpanzee or the human, we see a certain pattern in the evolution of neuroanatomy. Simplest organisms like the hydra or a jellyfish possess a diffuse nerve net, and not a brain/spinal cord. Slightly

more evolved creatures possess chains of ganglia, which are clumps of neurons, but still no brain/spinal cord. In organisms that are still higher, the clumping process seems to progress further, resulting in large, unitary clumps of neurons which we identify with the brain and spinal cord. What drives this clumping? Organisms that are higher up on the evolutionary ladder do have larger nervous systems, which means more neurons. But where is the necessity for these neurons to form clump(s)? Why can't a chimpanzee carry on with a diffusive nerve net, or why is not a hydra a proud owner of a tiny brain and spinal cord?

We will begin by stating, in simple terms, what neuronal clumping gives to an organism. We will make the arguments more precise as we go along, and provide quantitative data to support the case. One thing that clumping gives is reduction in the length of the "wire" that connects neurons. As neurons come closer, the wire that connects them becomes shorter. But why is it important or even useful to have shorter wires connecting neurons? To answer that question, we need to reconsider the very purpose, the reason to be, of the nervous system.

The nervous system, first and foremost, is a high-speed communication system that puts different parts of the body in rapid contact with each other. Coordinated movement of body parts is impossible without a rapid communication network passing signals back forth between the moving parts. This can be appreciated even in the simplest of our movements. Imagine, for example, that you had just stepped on a sharp object. Your first response would be to withdraw your foot—the one that is hurt—from the object. But it is not as if the affected foot is making its own local, private response to the injury without the involvement of the rest of yourself. If you just reflexively withdrew your hurt leg, without simultaneously tightening the muscles of the other leg, you would lose balance and fall, adding to the preexisting misery. Therefore, a problem that arose, say, in your right toe, quickly engages the muscles of *both* the legs. Usually, the response goes farther than that. The eyes are also directed toward the source of trouble; the whole head might turn to aid the eye movement. Your entire vestibular system, the part of the nervous system that maintains your balance, will be engaged, as you lift your right foot off the sharp object and view the menace. The shock of the pain might trigger a cardiovascular response, mediated by the autonomous nervous system, a part of the nervous system that controls the activity of your internal organs, and soon you might feel your heart pounding. Thus, though the stimulus is local, the response is orchestrated at the level of the entire body. The above event typifies what the nervous system gives to an organism. The nervous system allows you to live and act as *one* piece.

A key element necessary to enable such whole body coordination is rapid communication. One way of doing it is to use wiring that allows fast conduction. Conduction velocity of signals that travel along neural wiring increases with increasing diameter of the wire. Thus faster signaling requires thicker cables. But the use of fat cables is not a very viable option since wiring then takes up more volume. In larger nervous systems, wiring already takes up more volume. In humans, 60% of the brain volume is taken up by the wire. This is understandable since wire grows faster than the number of neurons. Even a simple calculation, which, of course, depends on a slightly unrealistic assumption that every neuron is connected to every other neuron in the

nervous system, shows that the number of connections increases roughly as square of the number of neurons. If you connect  $n$  points such that every point is connected with every other, the number of connections you would end up with is  $n(n - 1)/2$ . Since wiring necessarily dominates the volume of larger nervous systems, taking the “thicker cable” route to achieve faster signaling is perhaps not the most optimal.

Another way to achieve faster communication is by reducing transmission delay, keeping the conduction velocities the same. This can be achieved by minimizing the wire length since shorter wire means smaller delays. This can be done by moving neurons around within the volume of the body of the organism, such that, while keeping all the connections intact, the total wire length is minimized.

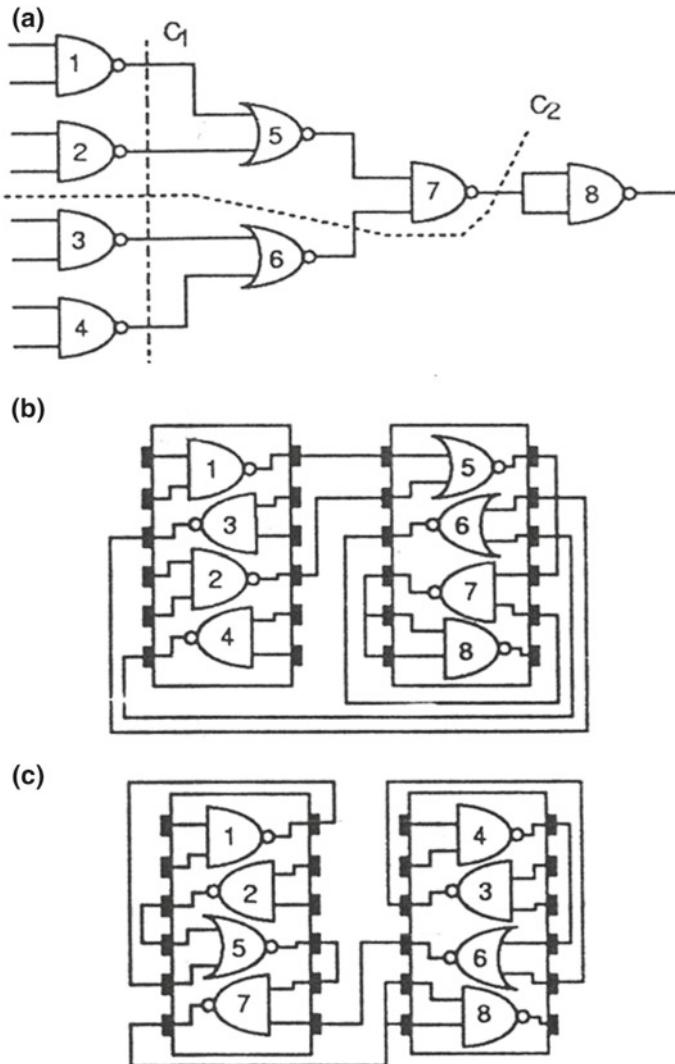
Interestingly, this problem of minimizing total wire length has an analog in the engineering world. Chip designers, who deal with the problem of planning layout of circuits in a chip, struggle to find ways of minimizing the total length of wire, a problem known as *component placement optimization*. Nature seems to be grappling with a very similar problem in designing the nervous systems of her creatures. Before considering the neural version of the problem further, let us make a quick detour into the engineering version of the problem.

## Component Placement Optimization

The goal of a chip designer is to design circuits that have certain *functional* and *layout* requirements. The functional aspect of the design refers to the function that the circuit is meant to perform: for example, add two 32-bit numbers or convert an analog signal into a digital signal. Such a circuit consists of several circuit elements like resistors, capacitors, transistors, or, at a slightly higher level, the logic gates (AND gate, OR gate, etc.). These circuit elements are connected by wire in specific ways, and are conveniently represented as *graphs*. Graphs are mathematical abstractions that can describe any system that consists of a network of connected units. The connections are referred to as “edges” and the points which are connected by the edges are known as “nodes” or “vertices.” A lot of real-world systems may be depicted as graphs, e.g., a computer network or a network of family and friends. Once the chip designers design the circuit—the circuit elements and their connectivity patterns—and ensure that the circuit functions as per the requirements, the functional part of the design is complete.

The subsequent phase, which consists of packing the circuit into individual chips and distributing the chips spatially on a circuit board, is known as *layout design*. The layout design phase consists of three subphases: (1) partitioning, in which a large circuit is partitioned into smaller modules, each of which is usually packed into a chip, (2) placement, in which the chips are placed at specific spatial locations on a circuit board and finally, (3) routing, in which the wiring that connects different chips are routed along the surface of the circuit board so as to satisfy a host of requirements.

An important objective that partitioning seeks to achieve is to minimize the total number of wires that run among various chips. This is because wires take up space



**Fig. 2.12** An illustration of partitioning. A circuit consisting of eight gates (a) is partitioned into two different ways (b and c)

on the circuit board and must be prevented from crossing over. The fewer the wires, the easier it would be to meet these constraints. The difficulty involved in solving the general partitioning problem is best illustrated with the simplest version of partitioning, viz., the bipartitioning problem which may be formulated as follows: how to divide a graph consisting of  $2n$  nodes, into two partitions each consisting of  $n$  neurons such that, the number of wires connecting the two partitions is minimized?

Figure 2.12 shows a few ways in which a graph consisting of eight nodes can be split into two partitions of four nodes each. A brute force search involves searching  $C_n^{2n} = (2n!)/(n!)^2$  different configurations. For  $n = 8$ , we have  $C_8^{2n} = (8!)/(4!)^2 = 70$ . But this number climbs very quickly as  $n$  increases. For example, for  $n = 100$ ,  $C_n^{2n}$  is a whopping  $10^{29}$ .

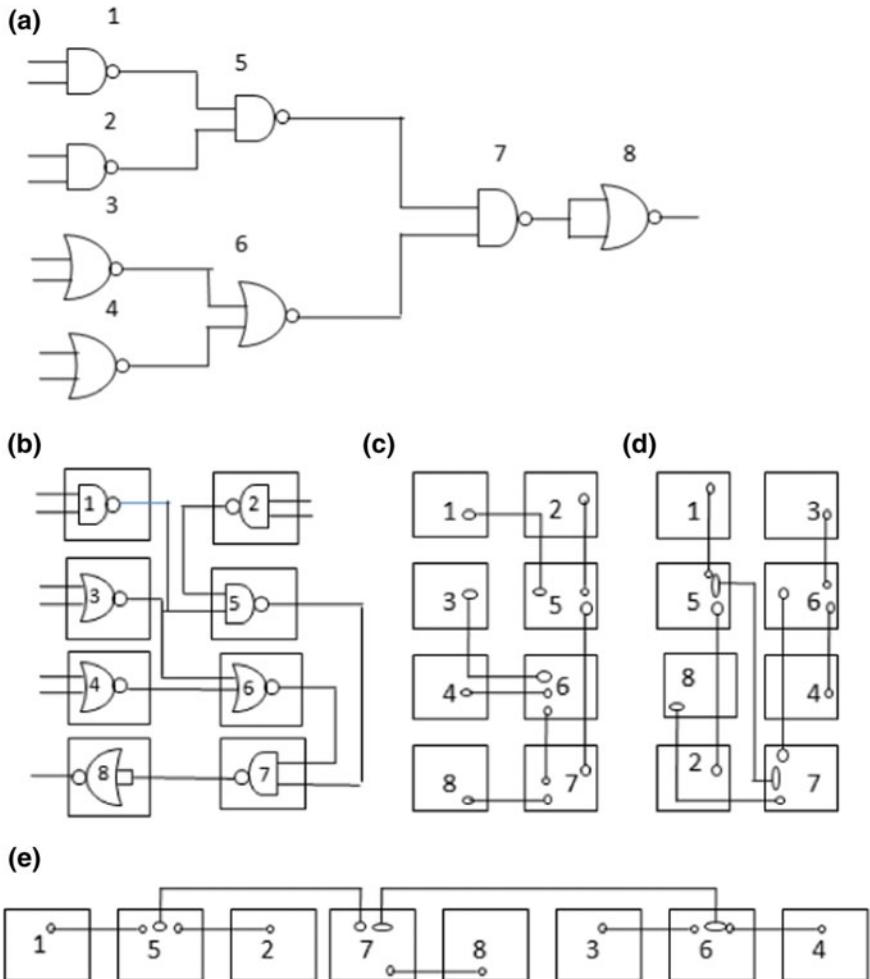
Note that, partition C1 which separates the circuit into {1, 2, 3, 4} and {5, 6, 7, 8} has four wires connecting the two partitions, while partition C2 which separates the circuit into {1, 2, 5, 7} and {3, 4, 6, 8} has only two wires running between the partitions. In a simple circuit like the above, it is possible to determine the optimal partitioning even by direct inspection. But complex optimization techniques are used to partition large circuits.

## Placement

Placement refers to the problem of positioning circuit components on the layout surface. A key constraint of placement is to minimize total wire length. We will consider with an example of how different placements of a circuit yield different wire lengths. Figure 2.13a shows a logic circuit with 8 gates, numbered from 1 to 8. The gates are placed on a rectangular grid structure shown in Fig. 2.13b. The “boxes” in which the gates are placed are spread out uniformly on the layout surface. The wire that connects the boxes also runs parallel to the axes of the layout surface. Figure 2.13c shows a symbolic representation, of the placement shown in Fig. 2.13b, in which only the boxes and the wire connecting them is displayed, while the contents of the boxes are omitted. Wire length between two adjacent boxes—horizontal or vertical—is assumed to equal 1. Thus, the length of the wire running between box 5 and box 7 in Fig. 2.13d is 3. By such calculations, the total wire length for placements in Fig. 2.13c, d equals 10 units. A linear placement shown in Fig. 2.13e, however, uses more wire (=12 units). Therefore, the placements of Fig. 2.13c or d, are preferred to that of Fig. 2.13e. In a simple example like the above, it is straightforward to discover optimal placements, but to find optimal placements in large, complex circuits with millions of gates, one must resort to sophisticated algorithms and adequate computing power.

## Routing

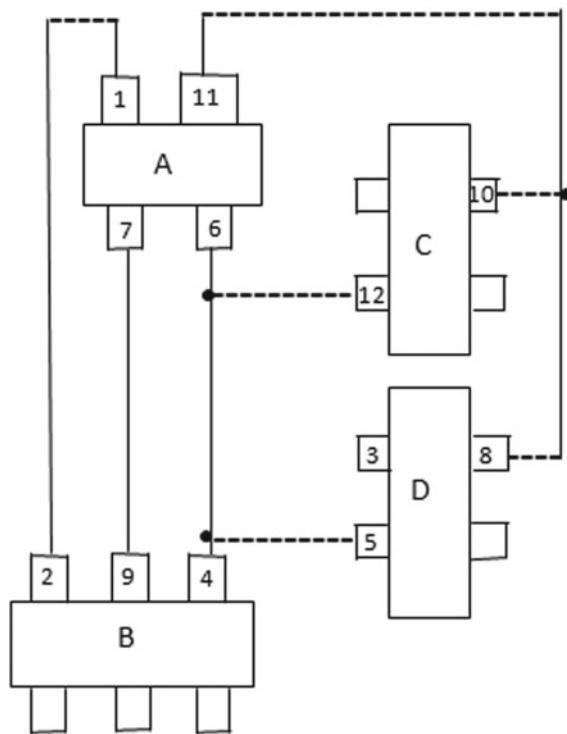
Once the circuit components are placed on the layout surface, they must be connected. Connecting the components while satisfying a variety of constraints is known as routing. In the placement example above, we assumed that the wire length is a simple function of the relative positions of the components that are connected. But additional constraints might make the situation more complicated. For example, one may have to make sure that there is sufficient separation between adjacent wires, or



**Fig. 2.13** An illustration of the placement problem

that the wires are of a minimum width, or that the wires do not crossover each other. A sample routing problem is illustrated in Fig. 2.14. The figure shows how four chips, whose ports numbered from 1 to 14 are connected. Note that, the connecting wires are not always the shortest possible wires due to additional constraints. The wires connecting ports 4 and 6 or ports 7 and 9 are the shortest possible. But the connection between ports 11 and 8 could have been made shorter, had we chosen an alternative path that starts from port 11, passes downward through the space between chips A and C, and then slips between chips C and D to reach port 8. But such a path has to crossover the wire connecting ports 6 and 12, and therefore forbidden.

**Fig. 2.14** An illustration of the routing problem

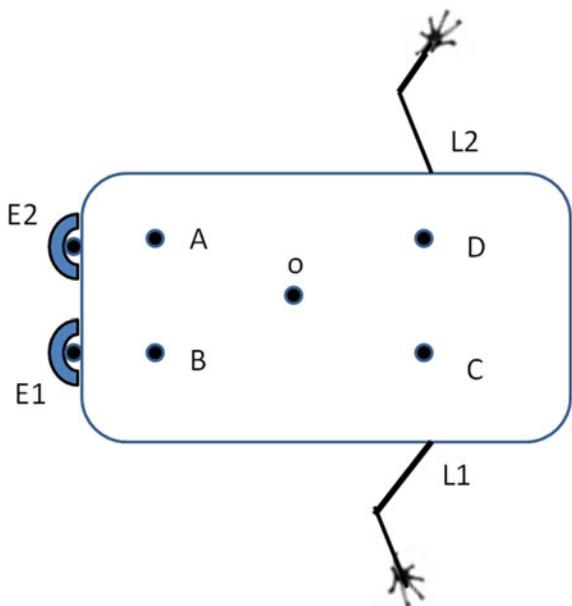


With this brief introduction to some of the key aspects of layout design, let us return to the problem of component placement in nervous systems.

## The Placement Problem in Neuroanatomy

We have seen above how, once the circuit is designed and the connectivity patterns are determined, the chip designers have to deal with the problem of spatially laying out the circuit. All the three stages of layout design—partitioning, placement, and routing—can be seen to have a common theme: minimizing the total wire length, while satisfying any other auxiliary conditions. Several constraints might influence layout choices. But minimizing the wire is the primary objective. This essentially simple yet powerful principle—termed the “save wire” principle, seems to shape the evolution of neuroanatomy through the aeons. The save wire principle is found to be capable of explaining the organization of the nervous system at several levels of hierarchy. (1) At the highest level, this principle explains the spatial location of the brain in the bodies of both invertebrates and vertebrates, (2) it explains the spatial layout of functionally related areas of cortex in higher animals, and also the location

**Fig. 2.15** A simple rectangular animal with two eyes (E1 and E2) and two legs (L1 and L2) and a nervous system consisting of a single neuron (N), which is located at several possible positions (A, B, C, D, or O)



of ganglia in invertebrates, and (3) at the lowest level, it also seems to explain the grouping (“partitioning”) and positioning (“placement”) of neurons in invertebrate ganglia.

Before poring over neuroanatomical data that support the save wire principle, let us consider some simple hypothetical nervous systems, apply the save wire principle to its organization, and see what kind of effects may be obtained. Let us begin with a simple, handcrafted organism. Let us imagine that the organism has a convenient rectangular body, with two eyes (E1 and E2) on either side in the front, and two legs (L1 and L2) on either side at the rear (Fig. 2.15). Assume that the nervous system of this animal has a single neuron, which receives single lines from either eyes and projects single fibers to the two legs. The question now is: where must the neuron be located so as to minimize the Total Wire Length (TWL) of the nervous system:

$$TWL = NE1 + NE2 + NL1 + NL2,$$

where NE1, NE2, NL1, and NL2 are the distances of the neuron N from E1, E2, L1, and L2, respectively. For a quick calculation, consider an imaginary, neat, “geometric” animal 10 cm long and 4 cm wide. Let the eyes be 1 cm away from the midline, on the anterior surface of the animal (Fig. 2.15). The legs are connected to the sides of the body 7 cm away from the “face.” Consider five possible positions of the neuron A, B, C, D, and O. A and B are off-center slightly toward the eyes; B and D are off-center toward the legs of the animal; O is closer to the centroid of the two eyes and the two legs. The total wire length with the neuron at each of the positions

**Table 2.3** Total wire length of the “nervous system” of the animal in Fig. 2.15

|   | Location of the neuron |    | Total wire length |
|---|------------------------|----|-------------------|
|   | X                      | Y  |                   |
| A | 1                      | 1  | 16.02             |
| B | 1                      | -1 | 16.02             |
| C | 5                      | 1  | 16.2              |
| D | 5                      | -1 | 16.2              |
| O | 2.3                    | 0  | 15.7              |

The single neuron in the nervous system is located at one of possible positions (A, B, C, D, or O) given by the coordinated ( $X$ ,  $Y$ ). Total wire length corresponding to each position is given in the last column

is shown in Table 2.3. Note that, the total wire length is minimized when the neuron is at a distance of 2.3 cm from the “face” on the lengthwise midline.

The above example is more an academic exercise and is obviously not a statement on the real nervous system. But even this trivial example can be made slightly more meaningful by adding certain realistic constraints.

To this end, we alter the number of fibers that run from the eyes to our solitary neuron. Assume each eye sends two fibers to the neuron, while the neuron sends only a single fiber each to the two legs. Naturally having multiple fibers is a happy step toward realism compared to the single fiber situation above. The human eye, for example, sends about a million fibers, via the optic nerve, to the brain. Similarly, the auditory nerve which carries sounds, coded as electrical signals, from ear to the brain, has about 30,000 fibers. But we are not yet ready to deal with real numbers. So, let us continue our studies of toy brains. With two fibers running from each eye to the solitary neuron, the total wire length is

$$\text{TWL} = 2\text{NE1} + 2\text{NE2} + \text{NL1} + \text{NL2}.$$

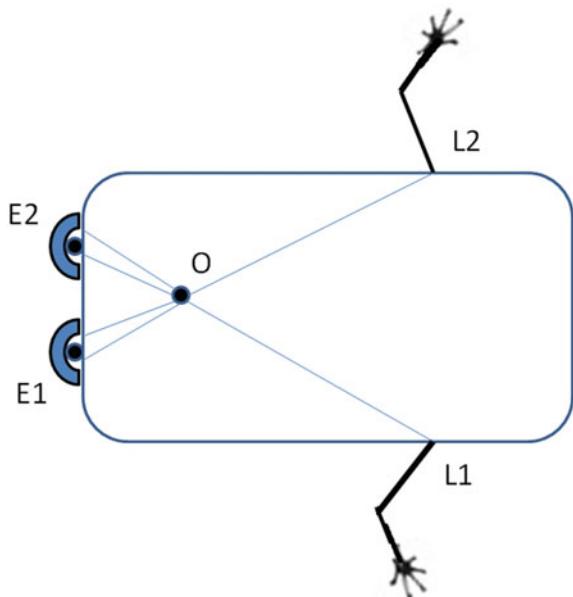
Since there is more wire toward the eyes, prudence has it that the solitary neuron is moved sufficiently toward the eyes. Our calculation shows that in this case, the ideal location of the neuron is at 0.5 cm from the “face” on the lengthwise midline (Fig. 2.16). If we let more fibers connect each eye to the neuron, the neuron moves even more toward the eyes, and farther away from the legs.

Now, it is the turn of the legs to bask in the limelight. Assume the neuron projects two fibers each to the legs, making the wire length

$$\text{TWL} = \text{NE1} + \text{NE2} + 2\text{NL1} + 2\text{NL2}.$$

In this case, the wire length is minimized when the neuron is 5.9 cm away from the “face” on the lengthwise midline. That is, the neuron is nudged closer to the legs in this case.

**Fig. 2.16** A simple “single-neuron” nervous system in which two fibers connect each eye to the neuron, but a single fiber connecting it to either leg



If we generalize the problem to a situation where each eye sends  $n_e$  fibers and each leg receives  $n_l$  fibers

$$\text{TWL} = n_e \text{NE1} + n_e \text{NE2} + n_l \text{NL1} + n_l \text{NL2}.$$

This time the neuron location is determined by the relative magnitudes of  $n_e$  and  $n_l$ . Taking one last step in our series of toy brains, imagine that there are  $m$  neurons, instead of a solitary one, all connected to the eyes and legs as we have just done. Since all the  $m$  neurons are connected in the same way to the eyes and legs, the optimal position of the single neuron in the previous case will apply to all the  $m$  neurons, which means that the  $m$  neurons are *clumped* together forming some sort of a “brain.” This time, the total wire length is

$$\text{TWL} = m(n_e \text{NE1} + n_e \text{NE2} + n_l \text{NL1} + n_l \text{NL2}).$$

Thus, the brain of our simple animal is best placed close to the eyes, if there are more fibers from the eyes, or pushed toward the legs if the number of brain-to-leg fibers dominates. This brings us a curious question with a serious neuroanatomical relevance. Where is our brain located? All the way up, behind the eyes, or way down near the legs? The question seems absurd and even silly since any right-minded person has a brain protectively encased in the cranial vault and not anywhere else. But a scientifically valid question is: why is the brain located in the head?

If you think seriously about this question, assuming you have not done so in the past, the first thing that flashes in your head—or the brain, wherever it is, if you do

not want to commit to its location as yet—is security. The brain is obviously safer tucked away in the boney cage of the skull rather than being deposited say in the abdomen along with the gut, liver and other viscera. But come to think of it, the head is not necessarily the safest spot in the body: its vulnerability lies at its base, in the fragile connection—the neck—with which it is linked to the rest of the body.

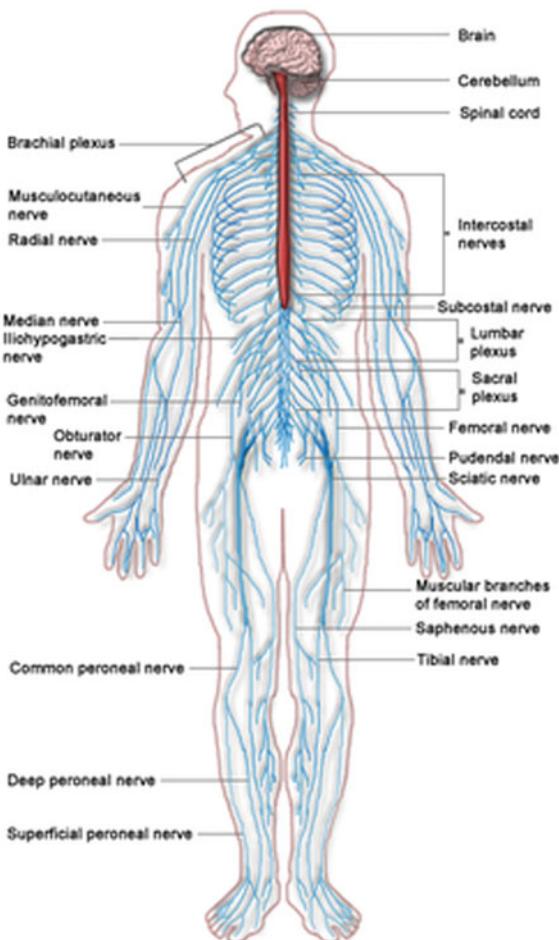
Interestingly, here too the save wire principle seems to come to our rescue in explaining the brain position. In our oversimplified example above, we saw how the ratio of the number of sensory fibers from the “front” or the anterior, and the number fibers going backwards or to the “posterior” seems to determine the position of the brain. We saw that if this ratio, called the anterior-posterior connection ratio, exceeds 1, the brain must be closer to the front of the animal. For the sake of precision, the “front” is defined as the part of the animal’s body that comes into contact with the external world as the animal moves around. Therefore, the front depends primarily on the animal’s normal heading direction.

In humans, the anterior-posterior connection ratio is found to exceed 1 by a clear margin, from a simple counting study of nerve fibers to/fro the brain and the spinal cord performed by Christopher Chermak and colleagues. To explain how this counting is performed, we must consider briefly the overall layout of human neuroanatomy. The human nervous system consists of a large clump of neurons centrally located, which is connected to the peripheral parts of the body through nerve fibers. The central clump is known as the central nervous system, while the nerve fibers that travel outside this central region constitute the peripheral nervous system. The central nervous system has two main subregions: the bulging portion at the top/front, the brain; and the long tail running backwards/downwards, the spinal cord (Fig. 2.17). Fibers that travel to/from the central nervous system are aggregated into bundles called nerves. Nerves that originate in the brain proper are known as the cranial nerves (there are 12 of them), and nerves that arise from the spinal cord are the spinal nerves (there are 33 pairs of them). The cranial nerves carry various sorts of sensory information from the sensory organs like eyes, ears, and nose to the brain and carry motor commands that control muscles of the face, or jaw movements. Thus, the domain of the cranial nerves is mostly confined to the head and neck region. Spinal nerves carry sensory information (mostly tactile) from the rest of the body; they also carry motor commands from the spinal cord to produce bodily movements.

Therefore, the fibers constituting the cranial nerves are the anterior connections, while the fibers of the spinal nerves form the posterior connections. Table 2.4 shows the number of fibers in each of the 12 cranial nerves. The fibers of the spinal nerves are not given explicitly, however, and are grouped broadly into dorsal (those going to the back) and ventral (those going to the front) fibers. It can be easily seen that the total number of anterior connections (12,599,000) far exceeds the number of posterior connections (2,400,000) with a ratio of 5.25. Thus, the anterior/posterior ratio is much greater than 1 in humans, which allows us to apply the save wire principle to argue strongly why the brain is located in the head in humans.

We next ask whether the same is true about the anterior/posterior ratio of animals. Unfortunately, the only organism, other than *Homo sapiens*, for which we have sufficiently detailed connectivity information to answer this question, is a worm known

**Fig. 2.17** The brain and the spinal cord form the central nervous system. The nerves that originate from the brain and spinal cord and are distributed over the entire body constitute the peripheral nervous system



as *C. Elegans*. More popularly called the nematode, this 1.3 mm long creature lives in soil. This organism has a nervous system consisting of exactly 302 neurons, a highly convenient figure for performing the kind of counting we have been preoccupied with. Moreover, the connectivity information of this 302-neuron nervous system is almost completely understood. The topology of this nervous system does not strictly come under chains-of-ganglia type, like those seen in the cockroach, for example, since there is also a cord-like structure known as the ventral cord. Nor can it be classified as brain-cord type, since there is no single mass of neurons that may be described as the brain, but only several groups of neurons concentrated around the pharynx. Anatomists identify 11 such neuronal groups, the ganglia, and have given them suitable names: (1) pharynx ganglion, (2) anterior ganglion, (3) ring ganglion, (4) dorsal ganglion, (5) lateral ganglion, (6) ventral ganglion, (7) retro-vesicular

**Table 2.4** Anterior/posterior connections in humans

| Anteroposterior connection ratio for the human brain |                     |
|--|---------------------|
| Cranial nerves                                       | Fibers (Both sides) |
| Olfactory  | ~10,000,000         |
| Optic  | 2,000,000           |
| Oculomotor   | 60,000              |
| Trochlear  | 6000                |
| Trigeminal   | 300,000             |
| Abducens   | 14,000              |
| Facial   | 20,000              |
| Cochlear   | 60,000              |
| Vestibular   | 40,000              |
| Glossopharyngeal                                     | 7000                |
| Vagus  | 70,000              |
| Accessory  | 7000                |
| Hypoglossal  | 15,000              |
| Total  | 12,599,000          |
| Spinal cord  |                     |
| Dorsal   | 2,000,000           |
| Ventral  | 400,000             |
| Total  | 2,400,000           |

Total number of fibers in cranial nerves are taken as anterior connections, while the number of fibers in spinal nerves are posterior connections. The anterior/posterior ratio is  $12,599,000/2,400,000 = 5.25$  (Cherniak 1994)

ganglion, (8) ventral cord, (9) preanal ganglion, (10) dorso-rectal ganglion, and (11) lumbar ganglion. Note that even the cord is treated as a ganglion in this scheme.

We will now consider the question of the brain's location in the case of *C. Elegans*. Anatomical studies reveal that among the 11 ganglia the first 3 (pharynx, anterior, and the ring) ganglia receive anterior connections, whose number adds up to 146 connections. Similarly, posterior connections to the remaining ganglia add up to 96. Thus, we have an anterior/posterior ratio of 1.52, far less than the corresponding figure in humans. But that probably may be correlated to the fact that there is no well-formed brain in *C. Elegans* compared to humans, and mammals in general. It is noteworthy that the ganglia that receive predominant anterior connections are themselves located quite anteriorly and may be thought to constitute the *C. Elegans*' "brain", of what comes closest to it.

Next, we consider the more general problem of optimal location of not just the brain but every part of the nervous system. Just as we were able to rationalize the location of the brain, in humans in and in *C. Elegans*, will it be possible to conduct a similar exercise about the entire nervous system and explain its architecture using the save wire principle? Such an exercise may not be feasible in case of humans, at the

current state of understanding of connectivity patterns in the brain, but is certainly feasible in case of a smaller organism like the *C. Elegans*. The problem may be more precisely formulated as follows: what is the optimal configuration of the ganglia in *C. Elegans* that minimizes the total wire length? Does that optimum correspond to the actual configuration found in *C. Elegans*? This study, also performed by Cherniak (1994) considers all possible permutations of the ganglia in *C. Elegans* body. Since the worm has a body length:diameter ratio of 20:1, it is practically a linear creature. Therefore, the layout consists of a set of 11 ganglia located on a single linear axis. Configuration counting is performed as follows. The connectivity pattern is fixed, and a given ganglion is allowed to occupy any of 11 fixed positions. That is, the ganglia play some sort of “musical chairs” with each of the 11 ganglia trying to occupy one of the 11 available seats. That generates totally  $11!$  (pronounced “11 factorial” and equals  $1 \times 2 \times \dots \times 11 = 39,916,800$ ) permutations. The wire length corresponding to each of these several million permutations or configurations is calculated, and it turns out that the actual configuration of the *C. Elegans* nervous system is also the one that has the smallest wire length ( $87,803\text{ }\mu\text{m}$ ; a micron is one-millionth of a meter). The second-best layout is only  $60\text{ }\mu\text{m}$  (or less than 0.1%) longer than the best one, which implies that the actual layout is strongly optimized to minimize wire length.

Continuing on the above lines with fiber counting efforts in a variety of nervous systems, Cherniak repeatedly found evidence for minimization of wire length in the brain. For example, in cat’s visual cortex, as in rat’s olfactory cortex, it was discovered that cortical areas that are connected have a greater chance of being next to each other, perhaps to serve the most probable purpose of keeping the wire length minimum. There is also evidence that even at single neuronal level, the dendritic arbors approximate the minimum spanning tree, which refers to a tree structure that connects a set of points such that the total wire length used is minimized. These and more data on similar lines, project the “save wire principle” as an important governing principle of neuroanatomy.

Minimizing neuronal wire length, as observed before, has obvious advantages, the most important of them being the minimization of communication delays. More wire also implies more volume of the brain. With the cranial vault placing a strong constraint on volume, it is not difficult to understand the demand to minimize wiring. Wiring also adds to metabolic and developmental cost to the brain. The importance of minimizing wire, in the brain as well as computer chips, is expressed lucidly by Carver Mead, a world expert in integrated circuit design, in these words: “Economizing on wire is the single most important priority for both nerves and chips.”

## Smart Wiring

Thus, there is convincing data that there is a clear evolutionary pressure on nervous systems to minimize their total wire length. There are strong physical reasons that support this principle. Longer wires mean: longer delays, more attenuation in signal,

more energy spent in carrying signals, and more metabolic expense in laying out wire over longer distances during development. The minimum wire principle has been able to explain several features of neuroanatomy: why is brain located in the head, why are cortical areas arranged the way they are, why are the ganglia of *C. Elegans* located the way they are, and so on.

But even this powerful organizing principle does not seem to shed any light on the question that we posed at the beginning of this chapter. What is the neuroanatomical basis of intelligence? Are there special anatomical features that correlate with intelligence or cognitive ability? The possibility of correlation seems moot since the example we began with—Einstein’s brain—did not provide dramatic results. But, interestingly, even on the more esoteric subject of linking neuroanatomy with intelligence, and marrying neural matter with neural spirit, the minimal wire principle seems to be relevant, and there is evidence that neural wiring patterns are correlated with intelligence.

In recent years, Diffusion Tensor Imaging (DTI) emerged as a powerful sophisticated tool for studying wiring patterns in the brain. This form of imaging is a variation of Magnetic Resonance Imaging (MRI) which has been playing a tremendous role, ever since the pioneering studies on humans in 1977, in providing unprecedented views of the body’s and brain’s internal structure. MRI is a technique based on how certain atomic nuclei, with hydrogen as a prime example, respond to magnetic fields and radio pulses. The response of the protons (hydrogen nuclei) thus excited is in the form of radio pulses, which constitutes the MRI signal. The MRI image provides essentially the distribution of hydrogen, and therefore of water. However, the response of excited protons is also influenced by the neighboring tissue. It matters whether the proton is in a neighborhood of cerebrospinal fluid, or fat, or white matter, etc. The neighborhood influences certain aspects of the signal which is used to tune MRI signal selectively so as to display certain types of brain tissue more prominently.

DTI is a variation of MRI that depends on the fact that molecules in liquids, in the fluids of the brain, are in a state of random, drifting motion known as diffusion. For example, a water molecule on an average diffuses over a distance of about 10  $\mu\text{m}$  during a period of 50 ms, bumping into brain’s microscopic structures such as cell membranes, fibers, or even large molecules, as they move around. The direction and range of diffusion of water molecules at a given location in the brain can be used to detect the presence of wiring. In a part of the brain dominated by cell bodies, as in gray matter, water is likely to diffuse in all directions with equal facility. But in the white matter, which constitutes wiring, water can only diffuse along the fiber and not across it. This introduces a directional bias in the way water diffuses in a neighborhood dominated by wiring. Thus, DTI is able to provide valuable information about wiring patterns in the brain.

Using DTI as a tool for measuring wiring patterns, Li and colleagues conducted a pioneering study to investigate the relation between wiring patterns and intelligence. The study involved 79 volunteers who were tested on their IQs, and classified into General Intelligence (GI) and High Intelligence (HI) groups. A variety of structural properties of the brain were considered, and the one most relevant to our present discussion about the significance of wire length, is a parameter called  $L_p$ , known as

mean characteristic path length. The entire cerebral cortex was subdivided into 78 regions.  $L_p$  is the average wire length that connects pairs of these regions. The study found that  $L_p$  is inversely correlated with IQ. That is the high IQ group on an average had shorter wiring connecting the brain regions considered in the study.

But the above was only an isolated study and one must await a considerable number of confirmatory studies before a comprehensive statement on the relevance of minimal wiring to intelligence can be made. Furthermore, we must note that minimal wiring is not the complete story, though it explains a good deal about neuroanatomy. The study also considers other network properties, which we do not mention here, in order to stay close to the line of reasoning that is being developed over the last few pages. There are also a large number of other studies which look at structural aspects (“certain brain structures are larger in smarter people”), other than wire length. In fact, the Einstein’s brain example that we began the chapter with, is one such study. There are studies which consider functional aspects too (“certain brain areas are more active in smarter people”). We make no mention of these studies at this point.

There are two reasons behind this conscious omission. Armed with an impressive repertoire of technological wizardry, brain science today spews out an immense flood of correlational data: activation in such and such area is more when a person is trying to recall memories, or activation in another brain area is less when the person is depressed. Such studies often state that two properties or events related to the brain tend to co-occur but do not explain why. The situation in a “hard science” like physics is different. Here, there is often a rich theoretical and mathematical framework, founded on a small number of fundamental laws, which can explain a vast array of phenomena. One would very much like to lay hold of a small number of neural laws, the so-called “neural information processing principles” which can explain the wide spectrum of neural and mental phenomena. There are some laws with powerful explanatory power (the minimum wire principle is itself one such an example) but such laws are few and scarce. Our objective in this book, we repeat untiringly, is to demystify, which is best done by trying to explain a variety of phenomena using a small set of laws, rather than inundate the eager reader with a mass of correlational studies. This is one reason the minimum wire principle was given more prominence in the last several pages, while a number of studies that correlate specific structures with intelligence are omitted. Principles demystify, correlational studies do not.

There is another reason why we consciously chose to hold back from a much more elaborate discussion of the vast literature on the neuroanatomical substrates of intelligence. Brain is a complex, curious organ with many facets—neuroanatomy, neurochemistry, and neuroimaging, the list can get quite long. Each of the facets provides a window, represents one line of attack onto the subject. But since each of these “facets” is unmanageably vast, experts tend to give prominence to one of these, at the risk of ignoring others. If our objective is to study the basis of intelligence, a singular study of the brain’s anatomy, however deep and scholarly it may evolve into, is not going to do the trick. One must gain the essential insights provided by each of these lines of attack, and later blend them, if possible, in a single grand synthesis.

Since we have deliberated at length on issues concerned with structure, it is now time to consider brain's function, and the mechanisms that form the basis of neural function. This will be the subject matter of the following chapter.

## References

- Angier, N. (2011). So much more than plasma and poison. *The New York Times*. [http://www.nytimes.com/2011/06/07/science/07jellyfish.html?\\_r=2](http://www.nytimes.com/2011/06/07/science/07jellyfish.html?_r=2).
- Butler, A. (1996). *Comparative vertebrate neuroanatomy*. New York, NY: Wiley-Liss.
- Campbell, N. A., & Reece, J. B. (2005). *Biology* (7th ed., p. 654). London: Pearson Education.
- Cherniak, C. (1994). Component placement optimization in the brain. *The Journal of Neuroscience*, 14(4), 2418–2427.
- Davis, H. (1996). Underestimating the rat's intelligence. *Cognitive Brain Research*, 3, 291–298.
- Diamond, M. C., Scheibel, A. B., Murphy Jr, G. M., & Harvey, T. (1985). On the brain of a scientist: Albert Einstein. *Experimental neurology*, 88(1), 198–204.
- Emery, N. J., & Clayton, N. S. (2005). Evolution of the avian brain and intelligence. *Current Biology*, 15, R946–R950.
- Fields, R. D. (2011). The hidden brain. *Scientific American Mind*, 22(2), 52–59.
- Fouts, R. S., & Fouts, D. H. (1996). *Project Washoe FAQ*. Washington: Chimpanzee and Human Communication Institute, Central Washington University.
- Holloway, M. (1997). Profile: Jane Goodall—Gombe's famous primate. *Scientific American*, 277(4), 42–44.
- Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Kartén, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, 6, 151–159.
- Kuhlenbeck, H. (1973). *Central nervous system of vertebrates* (Vol. 3, Part II). New York, NY: Arnold-Backlin-Strasse.
- Li, Y., Liu, Y., Li, J., Qin, W., Li, K., Yu, C., et al. (2009). Brain anatomical network and intelligence. *PLoS Computational Biology*, 5(5).
- Macphail, E. (1982). *Brain and intelligence in vertebrates*. Oxford, England: Clarendon Press.
- Mather, J., & Anderson, R. C. What behavior can we expect of octopuses? *Cephalopod Articles*. <http://www.thecephalopodpage.org/behavior.php>.
- McCall, B. (2001). Rat dreams: Oh great—now we'll have rat psychologists. *Discover Magazine*. From the October 2001 issue; published online October 1, 2001. <http://discovermagazine.com/2001/oct/featrat>.
- McCrone, J. (1991). *The ape that spoke; Language and the evolution of the human mind*. New York: Avon Books.
- Mead, C. (1989). *Analog VLSI and neural systems* (p. 102). Boston: Addison-Wesley.
- Sait, S. M., & Youssef, H. (1996). *VLSI physical design automation: Theory and practice*. Singapore: World Scientific.
- Spangenberg, D. B., & Ham, R. G. (1960). The epidermal nerve net of hydra. *Journal of Experimental Zoology*, 143(2), 195–201.
- This website presents data that relates intelligence to anatomical metrics of the brain. <http://serendip.brynmawr.edu/bb/kinser/Int1.html>.
- Witelson, S. F., Kigar, D. L., & Harvey, T. (1999). The exceptional brain of Albert Einstein. *The Lancet*, 353, 2149–2153.

# Chapter 3

## The World at the Level of a Neuron



*Swiftly the brain becomes an enchanted loom, where millions of flashing shuttles weave a dissolving pattern-always a meaningful pattern-though never an abiding one.*

—Sir Charles Sherrington

In the last chapter, we have seen how the structure of the nervous system changes with the complexity of the organism. We have also seen how as the size of the nervous system grows, under the pressure of the “save wire” principle, the nervous system evolves from a diffuse-nerve-net type to a brain-and-cord variety. These more evolved architectures of the nervous system are associated in general with organisms with a greater range of capabilities. The correlation between brain structure and intelligence, in the final analysis, is a weak one. Therefore, we are compelled to look for more precise indicators of brain’s capabilities.

Looking at it from a different angle, we can see how intelligence and brain structure need not be strictly correlated since intelligence is a matter of function. Structure and function need not have a one-to-one relationship though there can be an overlap. Invoking the computer analogy (which we will use sparingly, since it can sometimes be misleading), it is like estimating the capabilities of a computer by measurements of its CPU’s chassis. What a computer can do, first of all, is determined by the general specifications of its hardware, but also, more importantly, by the software loaded in its hard disk. Therefore, to understand how intelligence is represented in the brain, we must first identify brain’s “software.”

Brain’s hardware consists of the structure, the wiring, the connectivity patterns and the rest, while the “software” is more difficult to define because there is no precise correspondence. It certainly has to do with what brain structures do, their activity or function. Our discussion of the brain in this book began with brain’s “hardware” because it is easier to explain, but a major focus of this book is to define and describe what brain’s “software” consists of.

Anatomy and insight—the two terms seem to be almost contradictory. Textbooks on neuroanatomy describe brain's structure in excruciating detail, which of course is a compulsory diet for a student of medicine, but do little to explain the function of those structures. But perhaps it would be somewhat judgmental to say so because the job of explaining brain function must be left to neurophysiology. Even if we invoke an important tenet of biology that “structure inspires function,” in case of the brain, structure can often mislead our understanding of function. The history of neuroscience is full of such wrong turns and blind alleys. Therefore, traditional wisdom urges you to steer clear of anatomy and anatomists if you want to get an insight into brain function. But there are exceptions to every rule. An extremely influential book called *Vehicles*, written by V. Braitenberg, a preeminent neuroanatomist, is one such an exception.

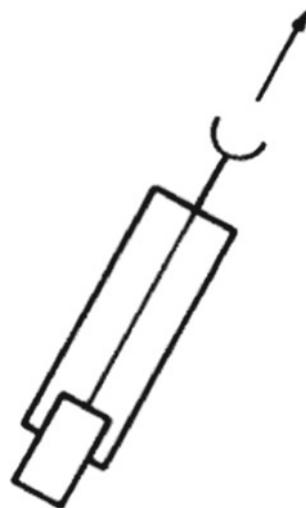
The complete title of this delightful book which reads “*Vehicles: Experiments in synthetic psychology*” makes the reader wonder what unearthly link might exist between vehicles and psychology. In this book, Braitenberg uses the word “vehicles” as a metaphor to an organism that possesses a nervous system. An organism is described as a vehicle that is capable of sensing the objects in the world around and navigate that world in ways that increase its survival. But for an organism to do all that, we expect it to have a complex and sophisticated nervous system. But this is where Braitenberg steps in, with his delightful little creations, to present an important insight which can be simply expressed as follows: complex behavior does not require a complex nervous system; a simple nervous system in its interaction with a complex environment can produce complex behavior. In order to illustrate this idea, Braitenberg presents a series of simple devices, of gradually increasing complexity, through various chapters of the book. These devices, the vehicles, are wired up to respond and move around in their environment in specific ways, displaying behaviors that resemble complex human emotions like love, fear, anger, and even shyness.

## Vehicles of Love and War

Braitenberg's vehicles are like little toy carts with wheels that children play with. Each vehicle has some sensors that measure properties of its ambience like temperature, light, humidity, etc. Signals measured by the sensors are fed to the motors that drive the wheels. Thus, environmental properties control the wheels of the vehicles and hence its movements.

Let us consider the simplest kind of these vehicles, the Vehicle-I, in which there is a single sensor in the front and a motor that drives a single wheel (Fig. 3.1). Assume that the sensor measures temperature and result of the measurement controls motor speed. The greater the temperature, the higher the motor speed. Such a vehicle speeds up in a hot environment and slows down in colder regions. Furthermore, on ideal, friction-free surfaces, it would follow a straight path, but the real-world frictional forces, between the surface and the wheels, make the vehicle deviate from its straight path. The vehicle will be seen to negotiate complex, winding trajectories, slowing

**Fig. 3.1** Braatenberg's Vehicle-I

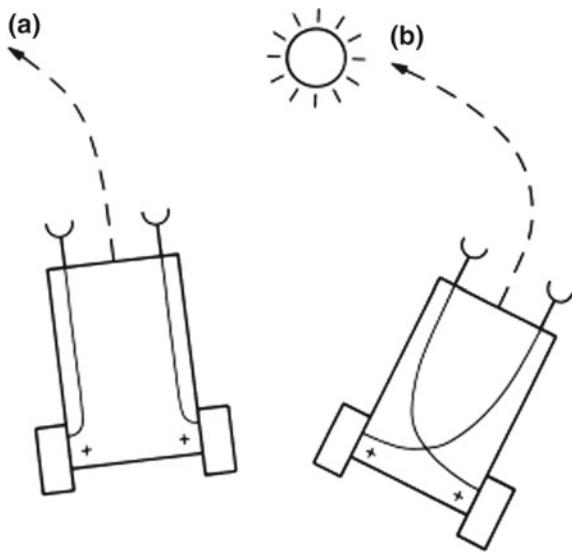


down, and speeding up, in response to ambient temperature. It would almost seem alive, following some complex inner law of life, while all along it was obeying a simple scalar, thermal life policy.

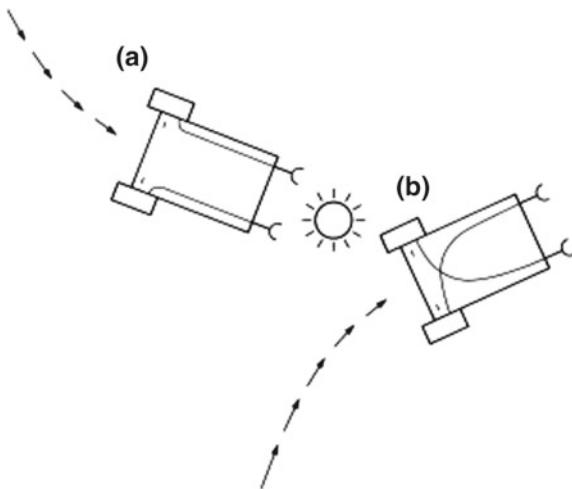
But Vehicle-I is too simplistic to be considered as a serious analog of a real-life organism with a nervous system, though it shows enough activation to be considered to possess life. Let us consider, therefore, the second class of Braatenberg's creations, the Vehicle-II. This class of vehicles has two sensors and two motors, one on each side of its rectangular body (Fig. 3.2). Consider the three possible architectures of Vehicle-II: (1) each sensor is connected to the motor on the same side only, (2) each sensor is connected to the motor on the opposite side only, (3) both sensors are connected to both motors. It is evident that the third case is simply a glorified form of Vehicle I. Therefore, we consider only cases 1 and 2. Assume that the sensors respond to light (it could be smell, or sound or heat or several other physical properties). The stronger the sensation the greater is the drive to the corresponding motor. In the case of Fig. 3.2a, if the light source is, say, to the right of the vehicle, the right sensor picks up a stronger signal than the left one, and the right wheel rotates faster. Thus, the vehicle will be seen to be running away from the light source. The opposite effect will be seen in the vehicle of Fig. 3.2b, since the wheel on the opposite side of the light source turns faster. In this case, the vehicle rushes towards the light source, increasing its speed as it approaches it closer and closer. Now let us imagine that these simple machines are housed inside real-looking creatures, soft and slimy. An unwitting observer of these vehicles would conclude that both the vehicles, first of all, dislike light, and express their dislike in contrary ways. The first one looks like a coward, fearful of light and its possible harmful effects. The second one hates light sources and rushes towards them aggressively as if to destroy them.

The configurations of the last two figures have only excitatory connections. That is, increased intensity of external stimuli can only increase the motor speed. Such an

**Fig. 3.2** Two variations of Braatenberg's Vehicle-II. **a** Sensors are connected to wheels/motors on the same side. **b** Sensors are connected to motors on the opposite side



**Fig. 3.3** Braatenberg's vehicle with negative connections. Two variants are shown: **a** one in which the sensors are connected to the motors on the same side, and **b** the other with connections on the opposite side



arrangement yields a limited range of behaviors. More interesting behaviors emerge if we introduce inhibitory connections, i.e., if we permit the motor to slow down when the corresponding sensor record increased intensity of stimulus. Figure 3.3 shows two variants: one in which the sensors are connected to the motors on the same side, and the other with connections on the opposite side.

In the vehicle of Fig. 3.3a, the motor on the “stimulus side” runs slower due to inhibitory connections. Therefore, this vehicle unlike its counterpart of Fig. 3.2a actually orients towards the stimulus. Contrarily, the vehicle of Fig. 3.3b turns away from the stimulus. But we may quickly note an important common difference between

the vehicles of Fig. 3.3 and those of Fig. 3.2. Both slow down when they approach close to the stimulus because the overall intensity of signal received from the stimulus increases with proximity, and the motors slow down. Here, the vehicle with “same side” connections simply approaches the stimulus and stops at a distance. This vehicle is influenced by two apparently contrary forces: one preventing it from coming too close to the stimulus, and the other preventing it from turning away from the stimulus. But the forces working on the vehicle of Fig. 3.3b are slightly different. While it may be able to draw too close to the stimulus, it is free and actually compelled to turn away from the same. Thus, vehicle of Fig. 3.3b displays a curious behavior. As it approaches a stimulus it slows down and once it is sufficiently close to the same, it gently veers away and goes off elsewhere!

We may describe the “feelings” of the vehicles of Fig. 3.3 as those of “love” since, unlike the vehicle of Fig. 3.2b, they do not make aggressive advances toward the stimulus. But this latter class of vehicles shows such rich shades of sophisticated “love.” The vehicle of Fig. 3.3a displays a quiet “adoration” of the stimulus, drawn towards it, but “shy” to draw too close. On the contrary, the vehicle of Fig. 3.3b shows a more fickle and fanciful love: as it approaches the stimulus it suddenly grows afraid of a “commitment,” changes its mind, and wanders away in search of other relationships!

Braitenberg’s vehicles are metaphors of real nervous systems. They show that to produce complex behavior, the organism need not be complex. It is the interaction of a simple organism with a complex environment that produces complex behavior. The vehicles have a common underlying theme: a set of sensors that respond to various environmental properties drive a set of motors through a network. All the subtle variations in the behavior can be seen to arise out of the network because it is the network that determines the relationship between the sensory input and motor output. Another important feature of the vehicles is that they are not “programmed.” There is a constant flow of information into the vehicle into its sensors to its motor organs over a network. All the “programming” the vehicle needs, or has, is encoded in the connections of the network. Here, we encounter a very important idea that the nervous system can be seen as a network of connections, between the sensory and motor organs, that determines the behavior of the organism.

The difference between the behaviors of the two vehicles above arises due to the nature of the connections—are the connections to the same side or opposite, are they positive or negative? The side to which connections are made may be classified as a structural property. But what are positive and negative connections in the real brain? To answer this question, we must begin our journey into brain’s function. We may begin by saying that neurons are not passive links between sensory and motor structures but active “processing units” that receive information from sensory areas, work on that information, and transmit the results to the motor areas. To understand this processing and transmission, we need to take a closer look at the neuron and its function. We shall look at a neuron as a complex electrical device, with electrical currents flowing, in rhythmic waves, all over its intricate arboreal body. We shall learn how neurons talk to each other by sprinkling minute quantities of chemical at each other. We shall learn about the complex electrical, chemical, and structural changes

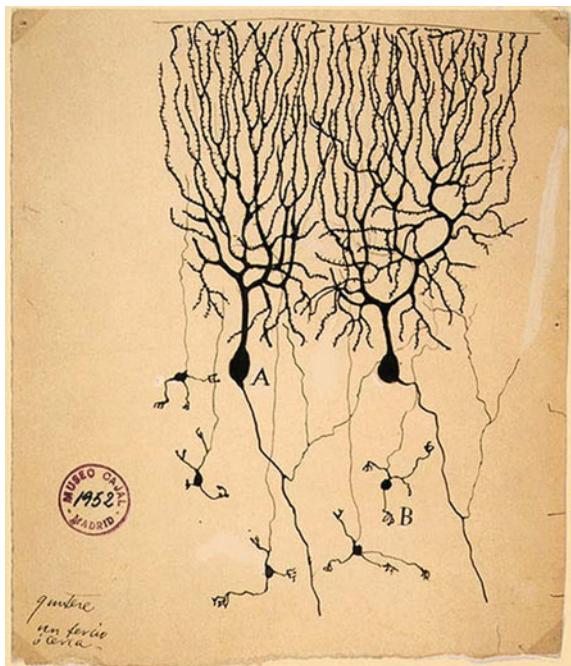
that occur in the microscopic world of brain tissue and how these changes support our brain's large-scale functions, creating our thoughts, emotions, and everything else that constitutes what we call our self.

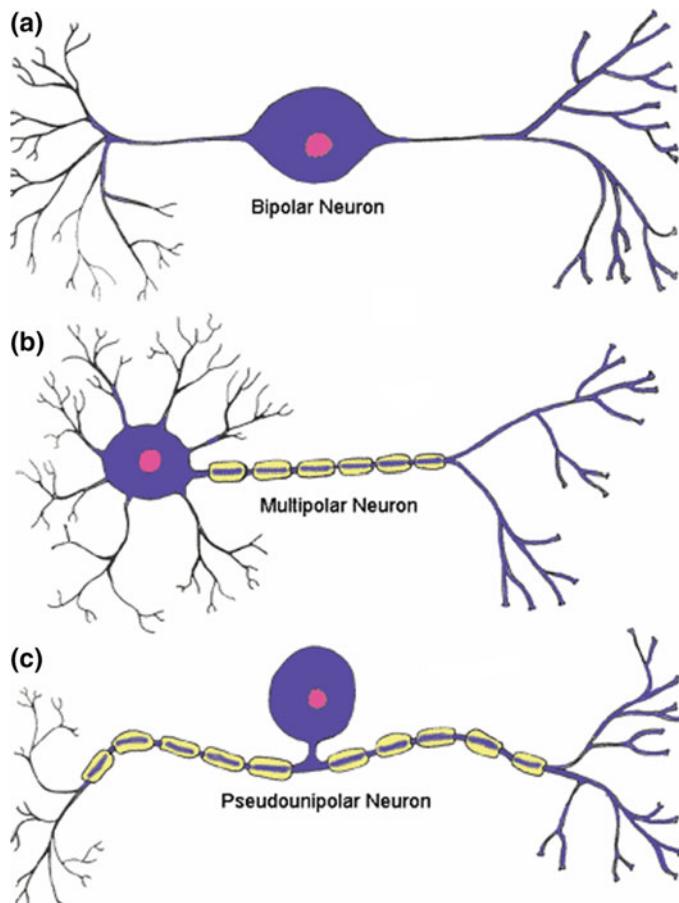
## The Neuron

A neuron, for all its arboreal abundance, is basically a cell. Like any other cell, it is a fluid-filled bag consisting of all the standard paraphernalia like the nucleus and nucleolus, Golgi bodies, mitochondria, a membrane that separates the rest of the world from itself and so on. But if a neuron is just like any other cell in the body, why aren't the other organs smart? Why is genius the special prerogative of the brain and not of gall bladder? There are indeed a few differences between neurons, the brain cells, and other cells of the body, which seem to make all the difference.

A distinctive feature of a neuron which can be noticed in micrographic pictures is the hairy projections that stick out of its cell body. Figure 3.4 shows a picture of a neuron, a specific type called the Purkinje neuron, drawn by Ramon y Cajal. It is an impressive instance of scientific art considering that it was hand-drawn in an era when microscopic pictures could not be photographed. The tiny spot in the middle of the neuron in the Fig. 3.4 is its cell body, formally known as the soma;

**Fig. 3.4** A Purkinje neuron drawn by neurobiologist Ramon y Cajal



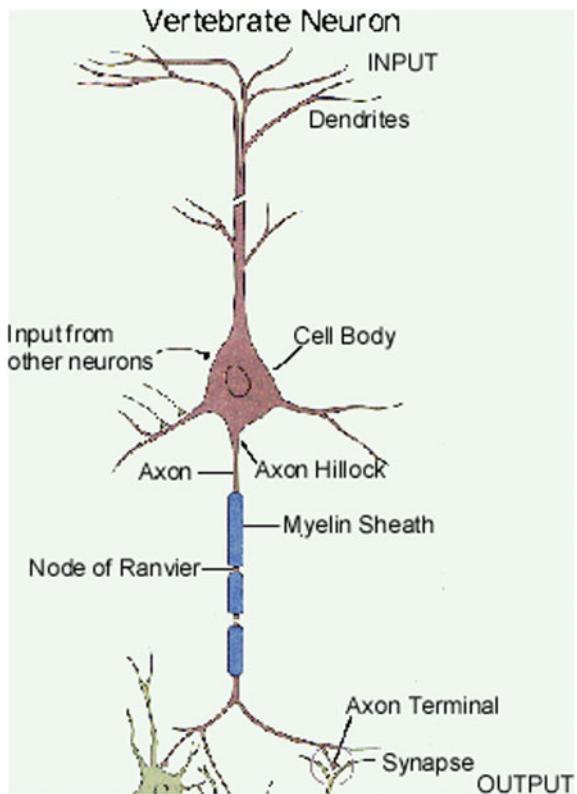


**Fig. 3.5** Three different neuronal morphologies. **a** a bipolar neuron, **b** a multipolar neuron and **c** a pseudounipolar neuron

the rest of its body is the “branches” or the “wire” of which deliberated at length in the previous chapter. The branches come in a variety of patterns—for example, thick bushy shrubs, or long slender stalks terminated by a short tuft—which produce a large variety of neuronal morphologies. For instance, compare the bipolar neuron (Fig. 3.5) found in the retina of the eye, with two single strands arising out of the soma and extending in opposite directions, with a Purkinje cell with its rich, arboreal outgrowth. The peculiar shapes of neuronal arboreal patterns often enable them to serve their unique functions.

A closer look at the branches of a neuron shows that they can be further segmented into two broad regions. Figure 3.6 shows a pyramidal cell, a ubiquitous type of neuron found in the brain, its name referring to its conical soma. Its branches can be seen to be distributed on either side of the soma. On one side, we notice many wires

**Fig. 3.6** A pyramidal neuron



emerging out of the soma, each of them branching repeatedly to form a dense arbor. These are called the dendrites, and the arbor formed by them, the dendritic tree. On the other side, we notice a single wire emerging from a slightly swollen part of the soma, known as the axon hillock. This single wire, known as the axon, extends to a distance before it branches into many *axon collaterals*.

The dendritic tree is smaller in size with a diameter of few hundred microns ( $1 \mu\text{m} = 1 \text{ millionth of a meter}$ ). The axons are typically much longer, in extreme cases extending to as much as a few feet. The axons are the neuron's long tentacles by which they reach out and make connections to each other. The axon is a neuron's mouthpiece with which a neuron sends out signals, in the form of bursts of electrical energy, to other neurons with which it is connected. At the point where one neuron meets another, the axon terminal of one neuron makes contact with the dendrite of another neuron. The meeting point between the axon of one neuron, and the dendrite of another, known as the synapse, occupies a very important place in all of brain's activity. Thanks to the synapse, and the myriad activities that take place within its narrow confines, a neuron is able to converse with another neuron.

The number of connections a single neuron can have to other neurons can vary. A typical neuron in the human brain receives about 1000 to 10,000 connections.

Some of the more densely connected neurons, like the Purkinje neurons, receive as many as 1 or 2 lakh ( $10^5$ ) connections. An adult brain has about 100 billion neurons. That makes the number of synapses about  $10^{11} \times 10^4 = 10^{15}$ , or a quadrillion synapses. In other words, brain is a network of 100 billion units with a quadrillion connections! To get an idea of the complexity of such a network let us compare it with the contemporary mobile network of the world. The number of mobile connections in the world had recently breached the barrier of 5 billion and is set to exceed 6 billion by 2012. Assume that each mobile has about 500 contacts in its address book, easily an overestimate, and therefore “connected” to so many other mobiles. That gives the mobile network about  $2.5 \times 10^{12}$ , or 2.5 trillion connections. The brain is definitely a much larger network but weighs only about 1.3 kg, all packed neatly in a volume of 14 cm × 16 cm × 9 cm.

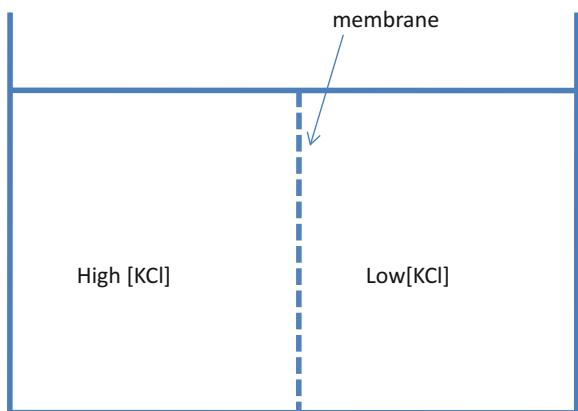
But that only gives us an idea only of the structural complexity of the network in the brain. The brain is not a static network. The connections among neurons make and break, even on the time scale of minutes, as we learn new things, and forget old ones. Furthermore, there are all the electrical and chemical signals that flash along the brain’s wiring system at speeds of hundreds of kilometers per hour, in our waking, as much as in our sleep. While structural complexity of the brain is impressive, it is the complexity of the signaling that drives brain’s function. The sources of human intelligence are mostly likely to be found in these signaling mechanisms. It is this functional aspect that has been ignored by the anatomical studies of Einstein’s brain. In order to understand brain function, we must first understand the electrochemical basis of a neuron’s function.

## Electrochemistry of a Neuron

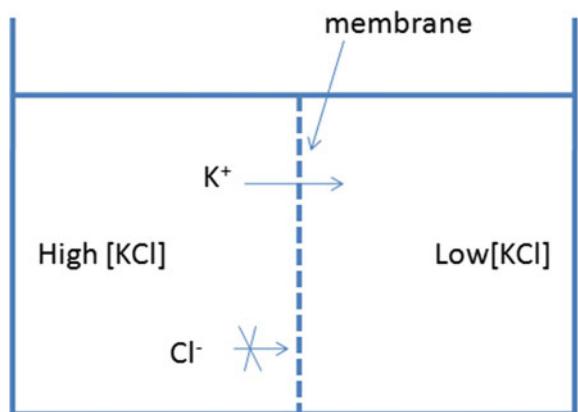
Imagine a beaker containing a salt solution like potassium chloride (KCl) (Fig. 3.7). The beaker has a central partition with the solution present on either side. Assume now that the compartment on the left has a higher concentration of KCl than the compartment on the right. Also, assume that the partition consists of a semipermeable membrane that allows only water to pass between the compartments. (A good example of such a membrane is the thin film on the inside of an egg, which allows only passage of water. When the egg is placed in pure, distilled water, water from outside enters the egg and the egg swells). By the familiar process of osmosis, water from the compartment with lower concentration of KCl now moves to the side with higher concentration, until the concentration in the two compartments equalizes, and that is not very interesting!

Now consider a slightly different kind of partition, a so-called semipermeable membrane that selectively allows only certain kinds of molecules (Fig. 3.8). KCl consists of an ionic medium filled with  $K^+$  and  $Cl^-$  ions. Assume that this membrane allows only  $K^+$  ions to pass between the compartments. Since there is a higher concentration of  $K^+$  on the left,  $K^+$  ions start moving from left to right.  $Cl^-$  ions too could have done the same but could not since the membrane blocks them. As  $K^+$

**Fig. 3.7** A beaker containing two compartments (left and right) of KCl, separated by a membrane permeable only to water



**Fig. 3.8** Beaker containing KCl in two compartments separated by a semipermeable membrane that allows passage only to  $K^+$



ions fill the right side of the compartment, there is a buildup of positive charges on the right side. These positive charges resist the further flow of positive charges ( $K^+$  ions) from left to right (since like charges repel). Thus, as  $K^+$  ions flow to the right, the initial chemical gradient becomes weaker and opposed by a growing electrical gradient. There comes a point when these two forces become equal. The flux of  $K^+$  ions then becomes zero and the solution reaches equilibrium.

Note that the solution in the two compartments now is *not* electrically neutral, which it was to start with. The compartment on the right has an additional number of  $K^+$  ions, and the one on the left has fewer. Therefore, voltage in the right compartment is higher than the voltage on the left side. Thus, there is net electrical driving force pushing  $K^+$  from the right to the left. But there is no net flux of  $K^+$  to the left at this point because  $K^+$  in the left compartment still continues to be at a higher concentration than that of the right compartment. Hence, there is chemical gradient from the left to the right acting on  $K^+$ . Thus at equilibrium, the left-to-right chemical gradient is exactly equal to the right-to-left electrical gradient.

Thus in this state of equilibrium—when the chemical gradient in the left–right direction is exactly opposed by the electrical gradient in the right–left direction, the right chamber is at a higher voltage than the left chamber. The voltage difference between the right and left chambers is related to the equilibrium concentrations  $[K^+]$  in the two chambers by the following formula:

$$V_1 - V_2 = \frac{RT}{Z_x F} \ln \frac{[X]_2}{[X]_1} \quad (3.1)$$

$V_1 - V_2$  Nernst potential for ion “X”

$[X]_1, [X]_2$  concentrations of “X”

$Z_x$  Valence of “X”

$R$  Ideal gas constant

$T$  Absolute temperature

$F$  Faradays' constant

$RT/F$  26 mV at  $T = 25^\circ\text{C}$  ( $Z_x = +1$ )

The above process of generating an electrical potential by creating a chemical gradient, made possible by a semipermeable membrane, also happens to be closely related to the process by which a normal battery works. Similar processes at work in a neuron produce a voltage difference between the interior and exterior of the cell. The neuron is permeated both within and without by a bath of ionic medium. Sodium and potassium ions are the key players in this medium, with chloride, calcium, and other ions playing more specific roles. Special structures embedded within the membrane that surrounds a neuron make the membrane semipermeable. These structures, known as ion channels, are pores within the cell membrane that are selectively permeable to specific types of ions. For example, sodium channels allow selective passage to sodium ions, and potassium channels to potassium ions and so on. This makes the situation in a neuron a bit more complicated than the simple situation considered above where the membrane is permeable only to one kind of ion, namely, potassium.

However, for purely pedantic reasons, let us consider a membrane endowed with only sodium channels. In normal conditions, sodium ion concentration is an order of magnitude higher outside the neuron than inside. Thus, we can imagine development of a positive voltage within in the cell with respect to the exterior. The sodium Nernst potential of a typical neuron turns out to be about 55 mV. Similarly, let us consider a membrane with only potassium channels. Normally, potassium concentration is much higher inside than outside, which should result in a negative voltage within the neuron compared to outside. The potassium Nernst potential is typically about –80 mV. But the situation becomes rather tricky when both kinds of channels exist and both kinds of ionic species (sodium and potassium) are present with the kind of distribution just specified.

With both kinds of channels present, we notice two opposing tendencies—one rendering the neuronal insides positive and the other negative. If only sodium channels were present, the membrane potential would equal sodium Nernst potential of +55 mV. Similarly, if only potassium channels were present it would be about

$-80$  mV. But with both kinds of channels present, the membrane potential takes a value between the extremes of  $+55$  mV and  $-80$  mV. What is that value? The answer can only be answered if we can assess the relative contributions of the two tendencies to final membrane voltage of the neuron. To make such an assessment, we must introduce two ideas—the idea that channels have conductance, and that they can be in OPEN or CLOSE states.

The ion channels are not permanent chinks in the neuron's membrane armor. They are well-regulated gateways through which material can flow in and out of the cell; they are the cell's windows onto the world. They can be in open or closed states, allowing certain molecules or blocking others. Open channels naturally have higher conductance, allowing easy passage to the ions that they are selective to. If  $E_{\text{Na}}$  and  $E_{\text{K}}$  are Nernst potentials, and  $g_{\text{Na}}$  and  $g_{\text{K}}$  are conductances of sodium and potassium ions respectively, then the neuron's membrane potential,  $V_m$ , may be expressed by the following easy formula:

$$V_m = \frac{g_{\text{Na}}E_{\text{Na}} + g_{\text{K}}E_{\text{K}}}{g_{\text{Na}} + g_{\text{K}}} \quad (3.2)$$

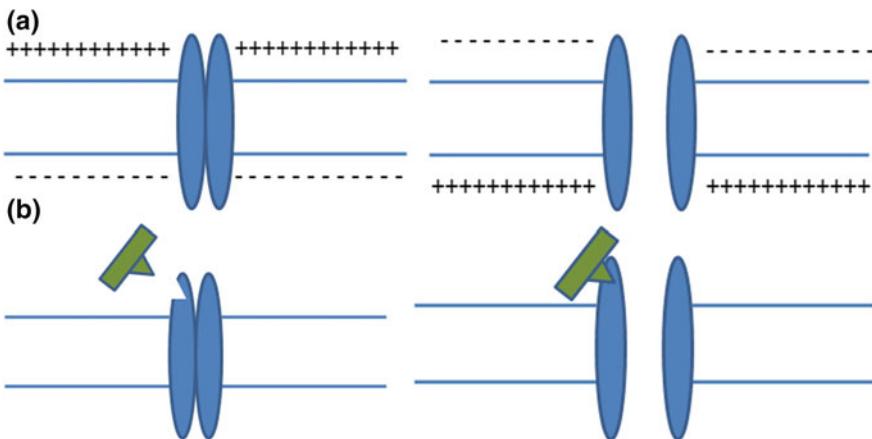
The above formula, it must be remembered, is a considerable approximation, and must only be considered as a pedagogic aid. Setting  $g_{\text{Na}} = 0$  (only potassium channels are present, or both channels are present but only potassium channels are open) or  $g_{\text{K}} = 0$  (only sodium channels are present, or both channels are present but only sodium channels are open), exclusively, we obtain  $V_m = E_{\text{K}}$  or  $E_{\text{Na}}$ , respectively, reinforcing the verbal arguments presented above. When both sodium and potassium channels are present and open ( $g_{\text{Na}}$  and  $g_{\text{K}}$  are both nonzero), the above formula tells us that  $V_m$  takes a value between a positive ( $E_{\text{Na}}$ ) and a negative ( $E_{\text{K}}$ ) extreme. Thus, the conductances  $g_{\text{Na}}$  and  $g_{\text{K}}$  may be viewed as two knobs which can be alternately turned up and down to increase or decrease membrane voltage. Opening/closing of sodium and potassium channels produce variations in membrane voltage which constitute the “activity” of the neuron, the constant chatter by which it makes itself heard by other neurons to which it is connected. What are the mechanisms that control the opening/closing (more formally known as *gating*) of ion channels?

Many influences can open or close ion channels, but we shall consider only two of them that are most relevant to our present purpose: (1) ligand gating and (2) voltage gating.

**Ligand gating:** An ion channel may switch from a closed to an open state when a molecule, known as the *ligand*, attaches to a part of the ion channel (Fig. 3.9a). This part, known as the receptor, is selective to only certain class of molecules which “fit” precisely in the slot offered by the receptor, as often described, like a key in a lock. Note that the event of binding between a ligand and a receptor can also close an open channel.

**Voltage gating:** Changes in voltage difference across the neuronal membrane, in other words, across the ion channel, since the ion channel straddles the membrane, can also open a closed channel (or close an open channel) (Fig. 3.9b).

It is not difficult to see how the above mechanisms offer the fundamental machinery that supports the rich tapestry of interneuronal conversations.



**Fig. 3.9** Channel gating. **a** Voltage gating and **b** ligand gating

## The Explosive Neural Response

The manner in which a neuron responds to an electrical stimulus is quite special, idiosyncratic, and is what makes the neuron so useful. What a neuron wants to say takes the form of fluctuations in its membrane voltage, and what another neuron gets to hear is in the form of tiny doses of current injected by other neurons. There is a familiar way in which systems respond to stimuli. For example, as you increase the pressure on the pedal, the vehicle gains speed until it reaches a maximum. As you turn the volume knob on a speaker, the volume increases unless it becomes intolerable to your neighbor. Thus with gradually increasing stimulus strength, a lot of real-world systems show a gradually increasing strength of response. Under ideal conditions, or within a small range of operating conditions, the relation between stimulus ( $S$ ) and response ( $R$ ) could be a matter of straightforward proportionality:

$$R = C \times S \quad (3.3)$$

where  $C$  is some constant that defines the proportionality.

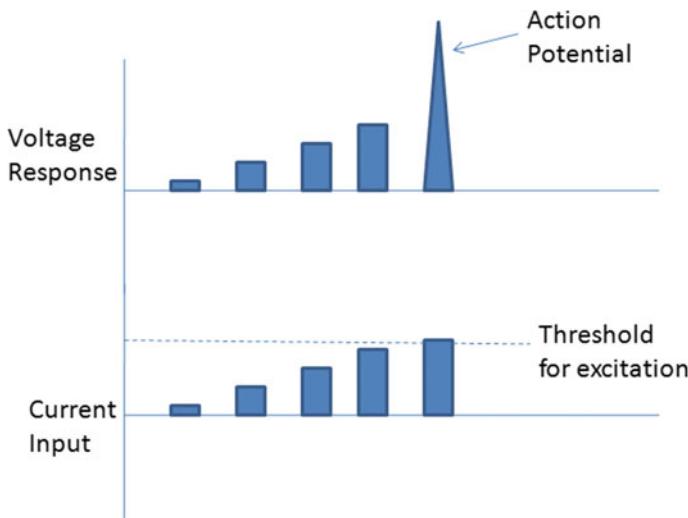
Systems in which the response ( $R$ ) is proportional to stimulus ( $S$ ) are known as linear systems because the plot of  $R$  versus  $S$  is a straight line. Study of linear systems forms a big part of physics and engineering, not exactly because real world is full of linear systems. Linearity is a mathematician's delight. Linear systems are easier to analyze. There is a large body of mathematics that has been developed over the last few centuries that can be brought to bear upon our study of physical systems—if they are proved to be linear. And when they are proved to be otherwise, *nonlinear*, that is—mathematicians often consider a small range of operation within which the system can be assumed to be linear (“a small part of a curve looks like a straight line”). There is even a rather pompous term for it: linearization. A great advantage

of the mathematical theory of linear systems is that it is universal. Solutions of linear systems are generic, applicable to linear real-world systems equally, irrespective of their physical underpinnings. Unlike linear systems, nonlinear systems cannot all be treated as a single large class. There are a large number of categories and subcategories of nonlinear systems, each posing its own peculiar difficulty to solution. These advantages of linearity often prejudice a mathematician to see linearity everywhere in the world.

But, thankfully, the world is not built to a mathematician's fancy. Nonlinearity abounds in nature. It crouches on the edges of linearity waiting to burst forth on the least occasion. The hallmark of a nonlinear system is a response that changes abruptly—almost explosively—as the stimulus is varied gradually. Examples of such a response too may be drawn from everyday experience. As the load is increased gradually, a rubber band stretches gradually but suddenly snaps when the load breaches a limit. With increasing current, an electric bulb glows brighter and brighter until the fuse blows at a critical current value. Phases of gradual buildup of strain energy between tectonic plates of the earth's interior are punctuated by sudden, explosive events when that pent-up energy is released in the form of earthquakes. But a nonlinear response need not always be destructive, like blowing fuses or natural disasters.

Nonlinearity can be constructive and most beneficial. The entire electronic and computer revolution has nonlinearity at its roots. The basic electronic components—diodes, transistors, operational amplifiers—are all nonlinear devices. Their strength lies in their nonlinearity. The basic logic gates—the OR, AND, and NOT gates—of Boolean logic that forms the fundamental mathematics of computer science are nonlinear. Therefore, it is quite interesting that a neuron's response to stimulus is nonlinear, and it is perhaps not surprising that the basic element—the neuron—in a computer of unparalleled power—the brain—is nonlinear.

A neuron's nonlinearity can be demonstrated by a simple experiment. Imagine a neuron to be a closed sphere, representing the neuron's cell body. Now if you inject a current into the neuron, its membrane voltage (always measured as difference between inside and the outside) increases, since there is a buildup of positive charge inside the neuron. If the current injection is stopped, and if the membrane were to be perfectly insulating, the new, raised membrane voltage would remain like that forever. (Students of physics may have noticed that a spherical neuron with an insulating membrane is a spherical capacitor, to a first approximation. But the presence of ion channels, particularly, voltage-sensitive channels complicates things.) But in a real neuron, the membrane has certain ion channels which are open all the time, the so-called *leakage* channels. Additional charge accumulated inside the neuron leaks out of these channels, bringing the membrane potential back to the baseline value, the resting potential. As the amplitude of the injected current is increased gradually, the upswing in the membrane potential also increases accordingly (Fig. 3.10). But when the current crosses a critical amplitude, the membrane potential raises steeply, disproportionately, to a sharp maximum and quickly falls back again to the baseline, as though a threshold has been crossed and the neuron is not the same anymore. This full-blown voltage spike produced by the neuron when the current injected crosses a threshold is known as the *action potential*.



**Fig. 3.10** A schematic depicting the all-or-none response of a neuron to a current pulse

The neuron's response is not very different from the behavior of a human being under increasing provocation: mild provocation elicits a mild response, but on increasing provocation, there comes a point when the individual displays an explosive, extreme reaction. This property of a neuron by which it responds explosively when the stimulus intensity exceeds a threshold is known as "excitability." It refers to the ability of the neuron to live in two very distinct states: the resting state of a steady, low membrane potential, and an excitable state of a transient, high membrane voltage. In addition to neural tissue, various types of muscle tissue also display this property. Henry Bowditch, an American physiologist who did pioneering work on the excitability of heart muscle notes that:

An induction shock produces a contraction or fails to do so according to its strength; if it does so at all, it produces the greatest contraction that can be produced by any strength of stimulus in the condition of the muscle at the time.

This binary response—a mild response versus a maximal one—is often referred to as the "all-or-none" response. The "none" part needs, of course, a slight modification: it refers not exactly to a lack of response, but more to a mild and subdued response.

How does the neuron produce this "all-or-none" response? The secret lies in the voltage-sensitive channels. The story of how the voltage-sensitive sodium and potassium channels work together in producing the neuronal action potential has been first worked out in the 1950s by two English physiologists A. L. Hodgkin and A. F. Huxley. In 1963, they were awarded the Nobel prize for their fundamental work in neuroscience.

## The Hodgkin–Huxley Experiments

To understand how the channels contribute to membrane voltage in general, and particularly to action potential generation, we must put these channels in perspective, in an appropriate framework. The simplest way to do so is to represent the neuron as an electrical circuit, with various relevant cellular components (mainly the ion channels and the membrane itself) represented as equivalent circuit elements. We have seen earlier that each ion channel has a conductance,  $g$ , which is higher when the channel is in open state, and smaller in closed state. We have also seen that each channel that is selectively permeable to a single type of ion (say sodium or potassium) has a voltage across itself. This voltage, known as the Nernst potential,  $E$ , depends on the ratio of the concentrations of the ion on either side of the channel. Therefore, when the membrane potential,  $V_m$ , is equal and opposite to the Nernst potential of a channel, no current flows through the channel. Any deviation of the membrane potential from the Nernst potential produces a current that obeys Ohms law of resistance. Thus, the current,  $I$ , through an ion channel of a given conductance and Nernst potential and for a given membrane potential may be given as

$$I = g(V_m - E) \quad (3.4)$$

In case of channels that are not voltage sensitive, the dependence of current on membrane voltage is straightforward and is a straight line. But in voltage-sensitive ion channels, the conductance depends on voltage,  $V_m$ , which makes the graph of  $I$  versus  $V_m$  more complicated. What would that graph look like? This was the question that Hodgkin and Huxley set out to answer.

One of the first things that Hodgkin and Huxley had learnt about the voltage-sensitive ion channels is that their conductance does not simply depend on the instantaneous value of the membrane potential. The conductance also depends on the recent values of  $V_m$ . That is,

$$\begin{aligned} g(t) &= f(V_m(t)) \text{ NOT TRUE} \\ g &= f(V_m(t), V_m(t - \Delta t), V_m(t - 2\Delta t), \dots) \end{aligned} \quad (3.5)$$

Therefore, in Eq. (3.4) above, when you find  $I$  varying with time, the variation could be driven by variation in  $V_m$  itself, or due to variation in  $g$ . In order to tease out the contribution of  $V_m$  to  $I$ , Hodgkin and Huxley used a clever method, known as the *voltage clamp*, to keep the  $V_m$  constant while the current is varying.

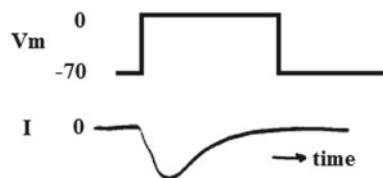
Hodgkin and Huxley performed such experiments individually on sodium and potassium channels. But to study the properties of one channel, the contribution from the other channel has to be annulled. This was accomplished as follows. When sodium channels were studied, potassium was taken out of the bath and potassium channels were blocked by use of drugs. Under such conditions, Fig. 3.11 shows the current flowing through sodium channels when a voltage pulse is applied to the membrane.

When the membrane potential is increased sharply from resting potential, sodium current appears to increase transiently before coming back to zero. Current flowing out of the neuron is considered positive by convention. Therefore, negative current indicates that sodium ions are flowing inwards. This occurs because, under resting conditions, as we are already familiar, there are more sodium ions outside than inside; these ions flow inwards when the sodium conductance increases. The transient increase in sodium current (irrespective of the sign) implies a transient increase in sodium conductance.

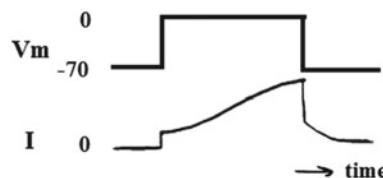
Similarly, the bath was made sodium-free and sodium channels were blocked when potassium channels were studied. Figure 3.12 shows potassium current in response to a membrane voltage shaped like a rectangular pulse. In this case, we note a more gradual, delayed increase in potassium current, which is depicted as a positive current because potassium ions flow inside out. This is because, as we have seen earlier, there are more potassium ions inside under resting conditions; these ions flow outwards when the conductance increases. Correspondingly potassium conductance also increases, and subsequently falls at a rate that is slower than that of the sodium channels. Hodgkin and Huxley also increased in steps the amplitude of the voltage pulse applied. They noticed that the corresponding amplitude of the variation in channel conductance also increased (Fig. 3.13).

The above experiments on the effect of membrane voltage on the conductances of voltage-sensitive channels can be summarized as follows:

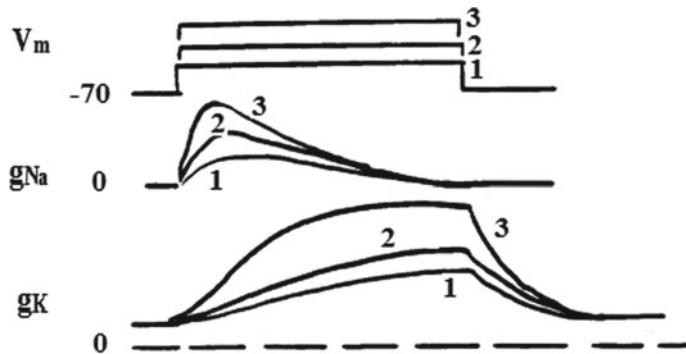
1. Ion channel conductance depends not just on instantaneous values of membrane voltage but also on recent history.
2. When the membrane voltage is sharply increased and held at the new value for a finite duration, channel conductance increases transiently before returning to baseline value. In case of sodium channels, the conductance rises sharply



**Fig. 3.11** Current in a sodium channel in response to a voltage pulse. Sodium current is negative which means that sodium ions flow inwards



**Fig. 3.12** Current in a potassium channel in response to a voltage pulse. Potassium current is positive which means that potassium ions flow outwards



**Fig. 3.13** Variation of  $g_{\text{Na}}$  and  $g_{\text{K}}$  in response to a pulse-like variation in membrane voltage

and falls rapidly towards the original value. In case of potassium channels, the conductance rises slowly and returns slowly towards the baseline value.

3. A greater increase in amplitude of membrane potential causes a larger transient in channel conductance.

If we keep these properties of voltage-sensitive sodium and potassium channels in mind, it is straightforward to understand how these channels are responsible for the generation of action potential. Earlier we have seen an oversimplified, approximate rule (Eq. 3.2) that relates the sodium and potassium conductances with membrane potential: when sodium conductance dominates potassium conductance, the membrane potential is close to  $E_{\text{Na}}$ , a positive potential. Similarly when potassium conductance dominates, the membrane potential approaches  $E_{\text{K}}$ , a negative potential.

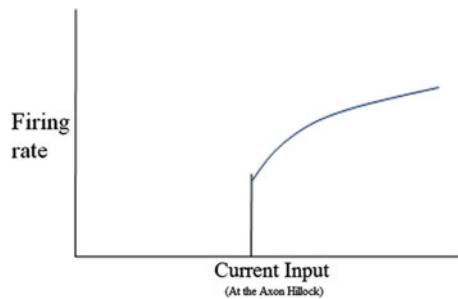
Now let us consider what happens if you try to increase membrane potential transiently by injecting a brief current pulse into a neuron. Sodium channels open quickly in response to the rise in membrane potential (as in Figs. 3.11 and 3.13). Opening sodium channels denote increased sodium conductance, which in turn pushes the membrane potential towards  $E_{\text{Na}}$ . Thus, we have here two processes that feed on and reinforce each other:

$$\text{Increased sodium conductance} \Leftrightarrow \text{increased membrane potential}$$

Such a system is known as a positive feedback system. Relevant quantities in a positive feedback system typically increase indefinitely until they reach their extreme values. This positive feedback is responsible for the rapid growth of membrane potential in the early phase of action potential, a phase that is dubbed the *rising phase*.

But the positive feedback relationship between sodium channels and membrane potential does not last long. Sodium channels which open quickly in response to the rising membrane potential also shut down quickly soon after (as in Fig. 3.13), a development that slows down the growth of membrane potential.

**Fig. 3.14** A step-like relationship between the current input to a neuron and its firing rate response



Another event that contributes to slowing down of the growth of the action potential is the slow opening of potassium channels (Fig. 3.13). We know that when potassium conductance dominates, the membrane potential approaches a negative value close to  $E_K$ .

Thus, after the initial rapid rise in membrane potential, the rapid closing of sodium channels, and the slow opening of potassium channels, result in the return of membrane potential from its maximum value (a positive voltage) to its original value the resting potential. The action potential thus generated has a fixed shape, amplitude, and duration not dependent on the input current pulse. The current pulse merely acts as a trigger; once the ball is set rolling, the subsequent evolution of the membrane potential waveform becomes independent of the initial trigger.

On stimulation by even stronger currents, a neuron generates not just one, but a series of action potentials. When the stimulation current is constant, and not stopped after a finite duration, neuronal firing too continues uninterrupted as long as the conditions in the bath (e.g., sodium and potassium ion concentrations) are maintained. As the current amplitude is increased further, the rate of firing also increases but saturates at a certain level of the input current. Thus a neuron's response in terms of its firing rate, to a constant current, has a step-like shape (Fig. 3.14). There is no firing up to a certain current level, beyond which there is an increasing firing rate which saturates. This step-like response, as we will see in the next chapter, forms the basis of one of the simplest mathematical models of a neuron.

That explains how a neuron generates its complex signals in the form of voltage spikes. But it does not tell us how these signals are conveyed from one neuron to another. Let us consider that story now.

## The Neuronal Handshake

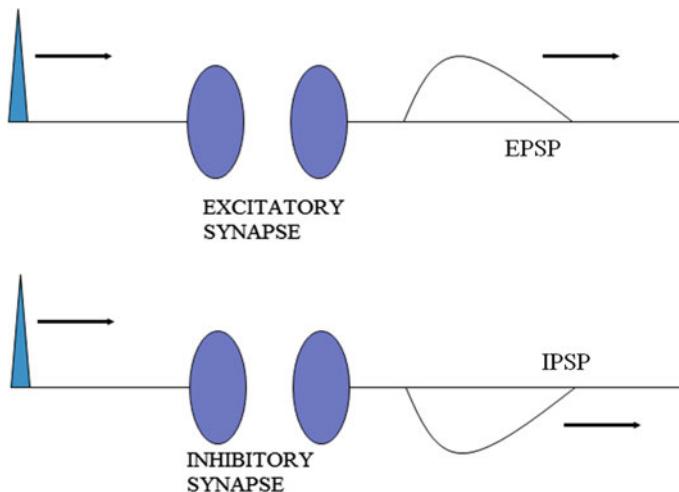
Since the time of Ramon y Cajal, it was known that neurons make connections with each other at special contact zones, later named *synapses* by Sherrington, an eminent English neurophysiologist. Cajal knew intuitively that neurons are independent, information processing units that communicate with each other over these connec-

tions, and that brain is an information processing network of neurons. But what he did not know was what exactly was transmitted from one neuron to another and how.

Sherrington did some pioneering work on spinal reflexes, a work which won him a Nobel prize in 1921. Reflexes are simple, rapid, automatic motor responses that are orchestrated at the level of spinal cord. Imagine yourself withdrawing your foot in alarm when you stepped on something sharp—a well-known form of reflex known as the withdrawal reflex. Your response will actually have two aspects: one visible and obvious, and the other invisible and covert. Rapid withdrawal of the affected foot is the obvious part of the reflex. While you were withdrawing one foot, in order to keep your balance, you were also necessarily making compensatory movements in the other foot such that your entire body can now be balanced on the unaffected foot. Thus it is clear that reflex consists of a whole pattern of activation of different muscles: some stiffen further while others relax. Sherrington guessed that this increase or decrease in muscle activation has a counterpart at the level of interneuronal communication. He predicted that there might be two kinds of synapses—the excitatory synapse, over which a neuron excites another neuron, and the inhibitory synapse, over which a neuron inhibits another.

Early evidence in support of Sherrington's insights was found in the studies of John Eccles, whose pioneering work on synaptic transmission brought him a Nobel prize. Eccles' group studied the synapse between the sensory neuron and the motor neuron in the spinal cord. Since a signal typically traverses a synapse in a unidirectional fashion, the neuron that sends a signal is called the presynaptic neuron, while the neuron that receives the signal is the postsynaptic neuron. Eccles and coworkers measured the voltage changes in the postsynaptic neuron in response to action potentials that arrive at the synapse from the presynaptic neuron. Surprisingly, they did not find action potentials on the postsynaptic side. Instead, they found slow, graded voltage waves. These waves are sometimes positive deviations from the resting potentials and sometimes they were negative. The positive deviations were named the Excitatory Postsynaptic Potentials (EPSPs) and the negative ones the Inhibitory Postsynaptic Potentials (IPSPs) (Fig. 3.15). Whether the responses are positive or negative seemed to have something to do with the nature of the synapse. But it was not clear what it is about the synapse that determines the kind of response produced in the postsynaptic neuron.

Knowledge of what exactly transpired between the presynaptic and postsynaptic terminals—within the narrow confines (now known to be 20 nm wide; 1 nm = 1 millionth of a millimeter) of the synaptic cleft—eluded neuroscientists for nearly a century. Since the early days of Ramon y Cajal, two rival views were prevailing about this matter. One view, the reticular theory, held that neurons in the brain are connected to form a syncytium; that is, neighboring neurons are connected by direct corridors such that their intracellular spaces form a continuum. Such a view seemed natural in the late nineteenth century, since microscopic observations of that time could not reveal any distinct gap between two neighboring neurons. An eminent proponent of the reticular view was Camillo Golgi who developed the famous staining technique which was the basis of microanatomical observations of neural tissue. The rival theory, known as neuron doctrine, was championed by Ramon y Cajal, who believed



**Fig. 3.15** Top figure: EPSP produced at an excitatory synapse. Bottom figure: IPSP produced at an inhibitory synapse

that neurons are distinct, isolated entities separated by a gap; neurons acted upon each other through mechanisms unknown at that time. The contrast between the two theories was brought forth quite dramatically in 1906 when both Golgi and Cajal shared the Nobel prize on the same podium.

The two rival theories of neural networks of the brain sought supporting evidence in two corresponding theories of synapse. The reticularists thought that synapses are solely electrical, where electrical signals are transmitted across the synapse by a direct conductive contact. The proponents of neuron doctrine believed that synaptic transmission is chemical, where information is conveyed by the presynaptic terminal to the postsynaptic terminal by emission of chemical signals. Interestingly, as early as 1846, German physiologist Emil DuBois-Reymond, discoverer of the action potential, proposed that synapses could be both chemical and electrical. But such an inclusive view was ignored due to the absence of conclusive evidence. Synapses were suspected to be chemical in nature because it was known for a long time that neurons responded to direct action of chemicals. Direct evidence in support of chemical synapses came with Otto Lewi's beautiful experiments with two hearts (see Chap. 1). Subsequent research resolved the tussle by showing that both kinds of synapses exist but with a predominance of chemical synapses.

Signal transmission at a chemical synapse represents a brief but complex chemical step in which an electrical signal (action potential) on the presynaptic side (the axon terminal) gets converted into another electrical signal (called the postsynaptic potential) on the postsynaptic side (the dendrite). The conversion from one electrical signal to another is mediated by a chemical process. There are three key chemical players involved in this conversion of the presynaptic electrical signal into a postsynaptic electrical signal. The first of these, the neurotransmitter, a molecule released from

the presynaptic terminal, is something like the presynaptic neuron saying “hi there!” to its postsynaptic neighbor. The postsynaptic neuron has molecular sensors called receptors, which can recognize the neurotransmitter and make appropriate responses to it. The receptor is associated with an ion channel, which can be in immediate vicinity or slightly removed from the receptor but located in the postsynaptic terminal. When the neurotransmitter is recognized by the receptor, the receptor sends a signal to the ion channel forcing it to open. It is this opening of the ion channel that generates the postsynaptic potential. Thus the neurotransmitter, the receptor, and the ion channel emerge as the three key players in the process of neurotransmission.

But why did Nature evolve such complex molecular machinery to construct a chemical synapse? An electrical synapse in which currents directly flow from the presynaptic to postsynaptic terminals would be far simpler to build. One feature that a chemical transmission brings, one that is absent in an electrical synapse, is *specificity*. The neurotransmitter and the receptor are often described respectively as the lock and the key of neurotransmission. The presynaptic terminal uses the key of the neurotransmitter to crack the lock of the receptor and open the gate of the ion channel. The gate opens only if the key is right for the lock. The messages of the presynaptic neuron in the form of outpourings of neurotransmitter will be heard only if the postsynaptic neuron has a receptor for it. Like in case of an electrical synapse, the chemical synapse is not an “always open” thoroughfare between two neurons. (Actually, even electrical synapses have some level of gating, but the chemical synapses have far stricter access than their electrical counterparts). Therefore, in case of chemical transmission, the presynaptic neuron is given a restricted access to the postsynaptic neuron.

Another important consequence of the complex molecular machinery (neurotransmitter, receptor and ion channel) of a chemical synapse is creation of two kinds of synapses, classified by the kind of effect the presynaptic neuron can have on the postsynaptic neuron. Life is much simpler with an electric synapse, in which the presynaptic terminal can only have more or less effect on the postsynaptic neuron depending on the conductance of the electric synapse. At a chemical synapse, the presynaptic terminal can have either a positive or a negative effect on the postsynaptic side, depending on the type of ion channel involved. If the binding event of the neurotransmitter and the receptor happens to open a postsynaptic sodium channel, for example, sodium ions from the extracellular space rush into the postsynaptic terminal thereby briefly increasing the local membrane potential. This positive deviation of the postsynaptic potential is called the Excitatory Postsynaptic Potential (EPSP) and therefore, a synapse in which neurotransmission opens postsynaptic sodium channels is known as an excitatory synapse. The story is very different when, instead of sodium channels, the binding event of neurotransmitter and receptor opens either potassium or chloride channels. When postsynaptic potassium channels open, potassium rushes out from within the terminal, thereby reducing the local membrane potential. Or when chloride channels open, negative charged chloride ions enter the terminal from outside, thereby reducing the membrane potential. Thus, in this case, there is a negative deviation in the postsynaptic membrane potential, which is known as Inhibitory Postsynaptic Potential (IPSP), and such synapses are known as

inhibitory synapses. Therefore, there are two types of chemical synapses, the excitatory or positive synapses, in which the presynaptic terminal excites the postsynaptic terminal, and the inhibitory or negative synapses, in which the presynaptic terminal inhibits the postsynaptic terminal.

It may be noted, therefore, that the postsynaptic potential, which can take positive or negative values, is fundamentally different from the action potential, which always consists of a positive deviation from the resting potential. Furthermore, we know that the action potential always has a fixed amplitude and duration, true to its all-or-none reputation. The postsynaptic potential is very different in this respect. Not only can it be positive or negative, its amplitude can take a nearly continuous range of values: not all-or-none but graded. The amplitude of the postsynaptic potential produced in response to a single action potential on the presynaptic side may be interpreted as a measure of the “strength” of the synapse. A synapse capable of producing a large amplitude postsynaptic potential may be described as a strong (excitatory or inhibitory) synapse, while a synapse producing a small amplitude postsynaptic potential is a weak synapse. We thus have the notion of the “strength” or “weight,”  $w$ , of a synapse, a quantity that can notionally take a range of values between a positive maximum and zero for excitatory synapses, and a negative minimum and zero for inhibitory synapses.

The idea that we can attribute a certain strength to a chemical synapse, and the fact that this strength can vary under the influence of experience, is perhaps one of the most important ideas of modern neuroscience. The pre- and postsynaptic terminals do not function as a single inseparable whole, but as two distinct entities with a variable, and tunable relationship. If the presynaptic neuron feels that it is not being given a fair hearing, it may decide to shout louder by releasing more neurotransmitter than usual. If a postsynaptic terminal decides to turn a deaf ear to the presynaptic ranting, it is free to do so by, say, reducing its receptor expression. Thus there are factors, both pre- and postsynaptic, that control the “strength” of the synaptic transmission. This changeability of synaptic strength is known as synaptic plasticity. An important tenet of modern neuroscience is that plasticity of synapses underlies a lot of learning and memory phenomena. When we learn a twenty-first century motor skill, like riding a bicycle with a mobile phone precariously pinched between the ear and the shoulder, or when we struggle to memorize those historic and not-so-historic dates, of wars and what not, on the night before an exam, we may be sure that a lot of synapses in our brain are frantically trying to readjust their strengths so as to encode the new information that is flowing in.

## The Neuron Sums It All Up

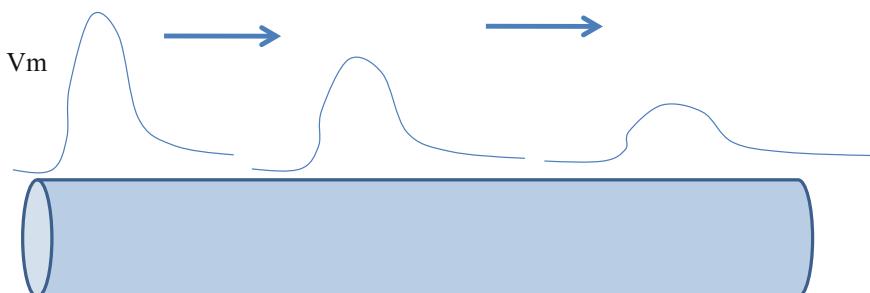
When a neuron transmits its signal across a synapse, the resulting PSP (EPSP or IPSP) produced on the postsynaptic side continues its journey on the dendritic branches of the postsynaptic neuron. In this journey, the voltage wave of PSP begins at the tip of a dendrite and winds its way through the dendritic branches towards the soma of

the neuron. Propagation of a voltage wave along a dendrite is not very different from propagation of an electrical signal along a telegraphic line. In fact, the equations that govern the two phenomena are identical. But the details of the underlying physics are different. The current in a copper wire constitutes electrons, while that in a neuron is ionic. There is also a great disparity in the velocities of propagation. Whereas propagation of an electrical signal in a metal wire is close to the speed of light, propagation in a dendrite is only a few meters per second. One of the characteristic features of wave propagation along a dendrite is that it is “lossy.” That is, as the wave moves down the dendrite it loses its amplitude and also spreads in time (Fig. 3.16).

Since the propagation is “lossy,” a single PSP originating from the tip of a dendrite might never be able to make it all the way to the soma. But usually, PSP does not come in isolation since APs do not come in isolation. When a volley of APs hit a dendritic terminal, the resulting PSPs buildup, wave upon wave, and might gather enough steam to last until the combined wave reaches the soma. This buildup occurring due to arrival of waves one after another in rapid succession is called *temporal summation*. When the wave thus built up at the soma is sufficiently large, it will be amplified, by the voltage-sensitive channels in the soma, into a full-blown AP.

There is another kind of summation that can cause wave buildup. APs may arrive at a large number of dendritic terminals, perhaps at different times, but timed in such a way that they reach the soma at about the same time. Thus, the tiny waves flowing from different remote branches of the dendritic tree come together at the soma creating a wave sufficiently large to trigger APs. This addition of waves arriving, not at different times, but from different points in space, is known as *spatial summation*.

It may be noted that the above division into temporal and spatial summation is artificial and pedagogic. In reality, both forms of summation operate together, and may therefore be described more aptly as *spatiotemporal summation*. Although we talk in intuitive terms about “buildup” of waves, it must be remembered that the waves can be both positive, if the cause is an EPSP, or negative if the cause is an IPSP. Thus, the positive and negative waves originating from different points of the dendritic tree, at different times, come together at the soma, augmenting (or annihilating) each other at the soma. At the end of all the addition and subtraction,



**Fig. 3.16** Lossy propagation of electrical signals in a dendritic cable. The voltage wave loses amplitude and also spreads in time

if a sufficiently large (positive) wave survives at the soma, the neuron gets excited and generates APs.

A curious analogy to this process of spatiotemporal summation may be offered. Let us assume that people in a large city like New Delhi plan to start a rally, demanding that the government must pass a certain bill. The rally is supposed to begin from a famous monument like India Gate located at the center of the city. People who intend to join the rally arrive from various remote corners of the city and its suburbs. They start from these several places of origin and proceed towards their trysting ground, the India Gate. Now, the situation at the India Gate suggests that it is in the best interests of the protesters-to-be, who are pouring from all over the city, to gather at the monument at the same time. The reason is: a battalion of the formidable city police is posted at the monument, ready to attack and disperse the protesting crowd. The only way the crowd can win is by outnumbering the police. Therefore, the optimal strategy for the protesters to gain upper hand is to make sure that the different streams of people flowing in from different directions, all arrive at the monument at the same time, or within a narrow time window. If different groups arrive one after another, each group will be dispersed by the police sooner than they arrive. And the rally will never take off. The analogy between the above situation and spatiotemporal summation occurring in a neuron is obvious.

A neuron basically receives influences (positive and negative) from other neurons, checks if their net effect exceeds a threshold level, and decides to fire or not fire. If the neuron decides to fire, the APs generated are broadcast to a number of other neurons. These other neurons, which, in turn, receive influences from yet other neurons, similarly make their own decisions to fire and spread the word. And the cycle continues.

We thus have a brain that serves as a stage for an incessant flux of neural signals, flowing from neuron to neuron, across various brain circuits, structures, and subsystems. Behind all that the brain does, inwardly in its thoughts and emotions, its sensations and intentions, and outwardly in speech, action and bodily control, there is nothing but this incessant neural activity, spreading from one structure to another, in measured rhythms. At the root of all this great ongoing cerebral drama, there is the solitary neuron that combines the decisions of several other neurons, with positive and negative “weightages,” checks if the sum crosses a threshold, and makes its own decision. An intriguing question that emerges at this point is: how does a network of such neurons perform the rich array of functions that the brain does? Answering this question is the motivation and subject matter of the following chapter.

## References

- Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, MA: MIT Press.  
Cannon, W. B. (1924). *Biographical memoir, Henry Pickering Bowditch, 1840–1911* (Vol. xvii, eighth memoir). Washington, D.C.: National Academy of Sciences.  
Fain, G. L. (2005). *Molecular and cellular physiology of neurons*. India: Prentice-Hall.

- Hodgkin, A., & Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117, 500–544.
- Johnston, D., & Wu, S. (1997). *Foundations of cellular neurophysiology* (Chapter 6). Cambridge, MA: MIT Press.

# Chapter 4

## Networks that Learn



*However opposed it may seem to the popular tendency to individualize the elements, I cannot abandon the idea of a unitary action of the nervous system.*

—Camillo Golgi, 1906.

### Why Neurons are not Logic Gates

Walter Pitts and Warren McCulloch had a challenging task ahead of them. They wanted to take a first shot at developing the mathematics of the brain. When faced with the unknown, it is natural to try to express it in terms of the known. McCulloch and Pitts knew something about the mathematics of the modern computer. They worked at the time of WWII. It was also the time when the first general-purpose electronic computer, the ENIAC, was built at the University of Pennsylvania. It performed computations a thousand times faster than the electromechanical computers that existed before. Most importantly, it could be programmed. The full power of the logic of computation, the Boolean logic, was at work in ENIAC. Popular media of those days described it as a “giant brain,” referring to its monstrous size.

Notwithstanding the media comments on the massive computer, it was quite tempting for McCulloch and Pitts to believe that brain is perhaps some sort of a compact, cold, and wet version of the ENIAC. Perhaps, the mathematics of the brain is akin to the mathematics that underlies the operations of ENIAC. But what exactly is the correspondence between the brain and the modern computer?

McCulloch and Pitts’ insight lies in noting that, thanks to its “all-or-none” response to stimuli, a neuron may be considered as a binary device. Its resting state may correspond to 0, while its excited state corresponds to 1. Furthermore, the researchers noted that a neuron is a thresholding device: only when the input exceeds a threshold, the neuron outputs a 1. Now the neuron’s input comes from a host of other neurons,

over a host of synapses, some excitatory and some inhibitory. Inputs over excitatory synapses push the target neuron toward excitation, while the inhibitory inputs prevent the target neuron from reaching that threshold. In simple mathematical terms, the net input to a neuron is expressed in terms of the inputs coming from other neurons as

$$\text{Net input} = w_1x_1 + w_2x_2 + \cdots + w_nx_n$$

where  $x_1, x_2, \dots, x_n$  are the inputs from other neurons and  $w_1, w_2, \dots, w_n$  denote the “strengths” (often also referred to as “weights”) of the corresponding synapses. The weights corresponding to excitatory synapses take positive values, while those corresponding to inhibitory ones are negative. If the net input exceeds a threshold, the neuron gets excited, and the output of the neuron,  $y$ , equals 1. If the neuron remains in its resting state, its output  $y$  equals 0. Therefore, the rules of operation of the neuron model proposed by McCulloch and Pitts can be summarized as follows:

$$\text{Net input} = w_1x_1 + w_2x_2 + \cdots + w_nx_n.$$

If(Net input > threshold) $y = 1$  “neuron is excited.”

If(Net input < threshold) $y = 0$  “neuron is in resting state.”

The great insight of McCulloch and Pitts lies in observing that a neuron which operates as a threshold device, can be used to perform logical operations. But for the above neuron model to perform logical operations, it must be possible to express logical operations using numbers, since the above neuron only deals in numbers. The recognition that logical operations can be expressed as numbers—binary numbers 0 and 1—has led to the development of a branch of mathematics known as Boolean algebra. Furthermore, the realization that not just logical operations, but a great variety of quantities can be expressed as 0’s and 1’s had created the foundations of the modern digital world.

A basic quantity in Boolean algebra is the logical variable. It denotes the truth of a statement. For example, let “ $x$ ” denote the truth of the statement: “the Ganga merges in the Bay of Bengal.” Since the statement is true,  $x = 1$ . Alternatively, the statement “the Ganga merges in the Caspian sea” is false and therefore the corresponding logical variable,  $x$ , equals 0.

Next, Boolean algebra introduces certain primitive logical operations that can operate on logical variables. There are 3 basic logical operations (there are more, but three should suffice for now), denoted by AND, OR and NOT.

The AND operation operates on two logical variables, say,  $x_1$  and  $x_2$ , and returns another logical variable,  $y$ , as the result.  $y$  is true only when *both*  $x_1$  and  $x_2$  are true. Since each logical variable takes two values (0/1), there are four combinations of values for  $x_1$  and  $x_2$ , each yielding a unique value of  $y$ , as follows:

If  $x_1$  is 0 AND  $x_2$  is 0,  $y$  is 0.

If  $x_1$  is 0 AND  $x_2$  is 1,  $y$  is 0.

If  $x_1$  is 1 AND  $x_2$  is 0,  $y$  is 0.

If  $x_1$  is 1 AND  $x_2$  is 1,  $y$  is 1.

Likewise, we may define the OR operation, which operates on two variables, say,  $x_1$  and  $x_2$ . In this case, the result,  $y$ , is true when *either*  $x_1$  or  $x_2$  is true. Again, we have four combinations of input values:

If  $x_1$  is 0 OR  $x_2$  is 0,  $y$  is 0.

If  $x_1$  is 0 OR  $x_2$  is 1,  $y$  is 1.

If  $x_1$  is 1 OR  $x_2$  is 0,  $y$  is 1.

If  $x_1$  is 1 OR  $x_2$  is 1,  $y$  is 1.

The last operation—the NOT—operates only on one logical variable,  $x$ , and returns the result,  $y$ . It simply denotes the logical negation of a truth and may be defined as follows:

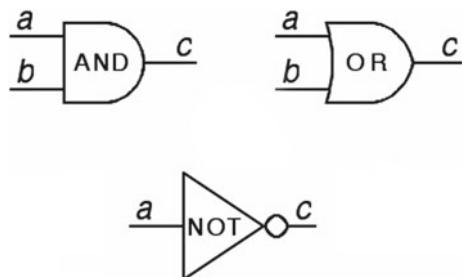
If  $x = 0$ ,  $y = 1$ .

If  $x = 1$ ,  $y = 0$ .

There are other logical operations like NOR, NAND, XOR, etc. But the above three—AND, OR, and NOT—are fundamental and complete in the sense that, more complex logical operations can be expressed as combinations of these basic three operations. Electrical engineers design circuits that perform these operations. Devices that perform logical operations are called logic gates. These gates are named after the logical operations they perform. The AND, OR, and NOT gates are pictorially represented as shown in Fig. 4.1. By connecting these gates in networks, it is possible to perform a large variety of computations, additions, multiplications, divisions, etc., which is the subject matter of Boolean algebra.

McCulloch and Pitts have shown how their neuron model can be made to behave like logic gates. For example, consider a neuron with two inputs and a single output. By judiciously choosing the weights (let  $w_1 = w_2 = 1$ ), and the threshold value ( $=0.5$ ), it is possible to make the neuron respond, say, like an OR gate. By similar choice of weights and threshold values, it is possible to design neurons that behave like an AND gate or a NOT gate. Once we construct these fundamental “logic neurons,” it is possible to connect them in networks and perform complex computations. The entire

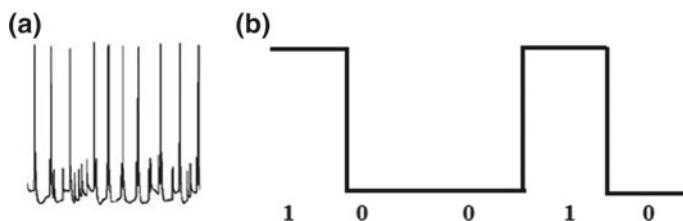
**Fig. 4.1** The basic logic gates



drama of digital logic and digital design can now be replayed in these networks of neurons.

The analogy between a neuron and a logic gate, the use of computer metaphor for the brain, is quite compelling. First, it can explain how the brain can represent a variety of things—sounds, smells, scenes, and somatic sensations—in the form of 0's and 1's, as mere flashes of neural spikes. It is puzzling how different parts of brain's surface known as the cortex with nearly identical histological architecture can represent diverse forms of sensory information. But by the application of computer metaphor, the issue is naturally resolved. Second, it can explain how brains can compute and perform logical operations. In one stroke, the McCulloch and Pitts' neuron model seems to offer immense possibilities for understanding the brain.

But the excitement did not last long. Soon discrepancies have begun to be noticed between the brain's design, as McCulloch and Pitts have conceived of it, and the digital computer design. The first prominent difference concerns the nature of the signal used by the two types of design. Brain uses a series of sharp voltage spikes known as action potentials, which we visited in the previous chapter (Fig. 4.2a). The computer uses rectangular waves (the “high rises represent 1's, while the low stretches represent 0's”) of fixed amplitude, resembling the wagons of a freight train, racing along the buses/wires of a computer (Fig. 4.2b). The second difference is about the clock speed or frequency of these signals. In a traditional computer design, all the pulses are at the same frequency, while in the brain the frequency of action potentials varies from neuron to neuron and with time. Thus, it is difficult to imagine that the code that underlies neural signals is a simple binary code of digital signals. Third, and most importantly, the brain learns from experience while every move of a computer has to be precisely and rigorously programmed. This last property—learnability—distinguishes the brain vastly from the computer. The McCulloch and Pitts neuron cannot learn. Its designers hard code its connections to make the neuron behave like a certain logic gate. A network of these neurons can perform, computer-like, complex operations. But such networks cannot learn those operations from “experience” on their own by adjusting their own connections. It seemed that unless such capability is incorporated, the direction of mathematical brain research indicated by the McCulloch and Pitts neuron, might very well be a sterile one.



**Fig. 4.2** Samples of **a** a neural signal and **b** a digital signal

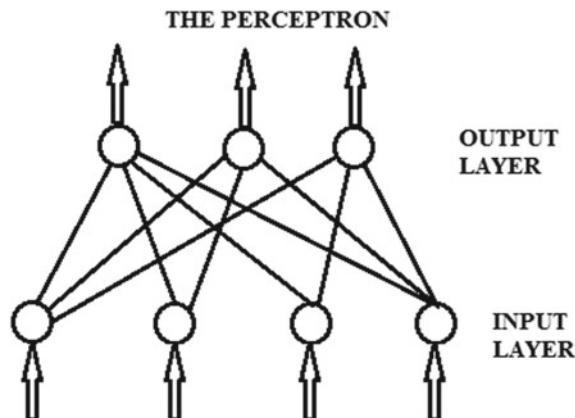
## Perceptrons

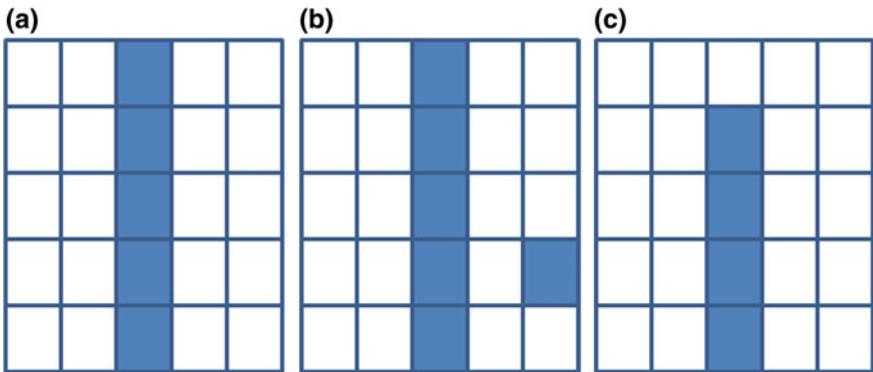
Perceptrons emerged to fill this vacuum. These were networks of neurons that can learn unlike the McCulloch and Pitts networks in which all the connections are fixed by a prior calculation. Perceptrons were developed in the '50s by Frank Rosenblatt, who was at Cornell University at that time. A bold ideal, namely, to discover "the fundamental laws of organization which are common to all information handling systems, machines and men included," was the motivation that drove Rosenblatt to create perceptrons. Like McCulloch and Pitts networks, perceptrons were originally developed to learn to recognize or "perceive" visual patterns, and hence the name. In fact, a perceptron is actually a network of McCulloch–Pitts neurons; what distinguishes them is the ability to learn patterns. A perceptron has two layers: an input layer which consists of the visual pattern to be recognized and an output layer consisting of a row or an array of McCulloch–Pitts neurons (Fig. 4.3). Neurons in output layer respond to and recognize patterns by producing a response that is unique to that pattern.

Consider, for example, a perceptron with a single McCulloch–Pitts neuron in the output layer. The input layer consists of a  $5 \times 5$  array of pixels that can represent a simple visual pattern like, say, a digit. Imagine that the single neuron in the output layer is trained to respond to the image of the digit "1" shown in Fig. 4.4, in which the third column in the input array is filled with 1's while the rest of the array has only 0's. Since the weights of the neuron,  $w_{11}, w_{12}, \dots$  are connected one-on-one to the input array, the weights can also be arranged as a  $5 \times 5$  array. Now consider a distribution of weights that is fashioned with a close resemblance to the input pattern. That is, all columns have -1's except column 3 which has 1's. Thus the response,  $y$ , of the neuron when the pattern in Fig. 4.4a is presented is given by

$$\begin{aligned} y &= 1, && \text{if net} > 0 \\ &= 0, && \text{otherwise,} \end{aligned}$$

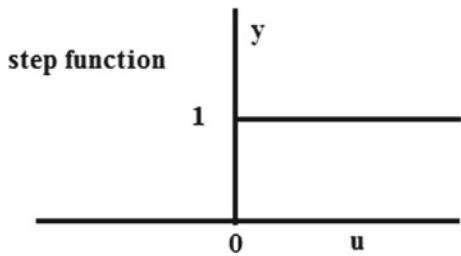
**Fig. 4.3** The perceptron.  
The neurons of the output  
layer are McCulloch–Pitts  
neurons





**Fig. 4.4** A  $5 \times 5$  array depicting the number “1,” in its full form (a) and its distorted versions (b) and (c)

**Fig. 4.5** Step function



where,

$$\text{net} = w_{11}x_{11} + w_{12}x_{12} + \dots + w_{55}x_{55} - b$$

For the mathematically inclined, the above rules may be expressed in a crisper mathematical notation as

$$y = \sigma\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij}x_{ij} - b\right), \quad (4.1)$$

where  $b$  is the threshold (also known as the bias), and  $\sigma(\cdot)$  is known as the step function (Fig. 4.5) defined as

$$\begin{aligned} \sigma(u) &= 1, & \text{for } u > 1, \\ &= 0, & \text{otherwise.} \end{aligned} \quad (4.2)$$

Our objective is now to construct a neuron which responds with a 1, when the pattern “1” in Fig. 4.4a is presented, and with a 0 for any other pattern. Since we have already chosen the weights ( $w_{11}, w_{12}, \dots$ ), now  $b$  remains to be chosen. The value of  $\text{net}$ , when pattern “1” of Fig. 4.4a is presented, can be easily calculated to be 5, by explicitly summing the terms in Eq. 4.1. Only the pixels in the third column contribute to the sum, since other input pixels are all 0’s. Let us now consider what happens when we present a distorted version of pattern “1” with extraneous, “noisy” pixels, or missing pixels in the body of “1” (third column). Consider the pattern in Fig. 4.4b which is a single extraneous pixel (its value is flipped from 0 to 1). In this case, the “ $\text{net}$ ” adds up to only 4, since one of the terms in the summation has now dropped to  $-1$  from 0. Similarly, consider a pattern in which one of the pixels in the body of “1” is set to 0 (Fig. 4.4c). Again “ $\text{net}$ ” adds up to only 4, since now one of the terms in the summation has dropped from 1 to 0. Thus when the input pattern corresponds to a perfect “1” (of Fig. 4.4a)  $\text{net}$  evaluates to 5, while with any deviation from the perfect case, the value of  $\text{net}$  is reduced. Now let us choose a value of the threshold,  $b$ , such that the neuron response,  $y$ , is 1 only for the perfect pattern. That choice is obviously any value that lies between 4 and 5. Thus, for  $b = 4.5$ , for example, the neuron responds with a 1, only for the perfect “1.” Similarly, it is possible to “design” neurons that respond to a “2,” a “9,” and so on. But that would still be a “designer” network and not one that is self-taught. We now introduce mechanisms by which a network of the kind described above can learn to respond to a set of patterns, by changing its weights in accordance to a set of sample patterns. In formulating such a learning mechanism lies the pioneering contribution of Frank Rosenblatt.

The way a perceptron can learn to respond correctly to a set of patterns can be best demonstrated using a simple example. Let us revisit the OR gate that we considered earlier in this chapter. We have seen that a McCulloch–Pitt’s neuron given as

$$y = \sigma(w_1x_1 + w_2x_2 - b)$$

where  $\sigma(\cdot)$  is again defined as

$$\begin{aligned} \sigma(u) &= 1, & \text{for } u > 0, \\ &= 0, & \text{otherwise,} \end{aligned}$$

and  $w_1 = w_2 = 1$ ,  $b = 0.5$ , can serve as an OR gate. We can easily verify this by substituting the four combinations of values of  $x_1$  and  $x_2$ , and check if the actual output,  $y$ , equals the desired output,  $d$ , of the OR gate as defined in Table 4.1.

But let us assume that neuron’s weights are now flawed and the neuron does not behave like an OR gate anymore. Take, for example,  $w_1 = 1$ ;  $w_2 = -1$ , and  $b = 1.5$ . The actual output,  $y$ , does not match the desired output,  $d$ , for three out of four patterns, as shown in Table 4.2.

**Table 4.1** Truth table for OR gate

| $x_1$ | $x_2$ | $d$ | $y$ |
|-------|-------|-----|-----|
| 0     | 0     | 0   | 0   |
| 0     | 1     | 1   | 1   |
| 1     | 0     | 1   | 1   |
| 1     | 1     | 1   | 1   |

**Table 4.2** Truth table for OR gate along with the actual responses for a neuron

| $x_1$ | $x_2$ | $d$ | $y$ |
|-------|-------|-----|-----|
| 0     | 0     | 0   | 0   |
| 0     | 1     | 1   | 0   |
| 1     | 0     | 1   | 0   |
| 1     | 1     | 1   | 0   |

We will now introduce a formula by which the perceptron's parameters ( $w_1$ ,  $w_2$ , and  $b$ ) are adjusted in small steps such that its actual output,  $y$ , equals the desired output,  $d$ , for all patterns. The formula may be expressed as follows:

$$w_1 \leftarrow w_1 + \eta(d - y)x_1. \quad (4.3a)$$

The arrow ( $\leftarrow$ ) symbol denotes a substitution of the quantity on the left of the arrow, by the quantity on the right. Here it means that  $w_1$  is replaced by  $w_1 + \eta(d - y)x_1$ . The term  $\eta(d - y)x_1$  denotes the correction to be given to  $w_1$  so as to bring the actual response,  $y$ , closer to the desired response,  $d$  ( $\eta$  is a proportionality factor, usually less than 1). Note that  $w_1$  gets actually modified only when  $y$  and  $d$  are different. Similar rules of correction are applied to the remaining parameters:  $w_2$  and  $b$ .

$$w_2 \leftarrow w_2 + \eta(d - y)x_2, \quad (4.3b)$$

$$b \leftarrow b - \eta(d - y). \quad (4.3c)$$

Let us try to understand in intuitive terms the meaning of the above equations. Consider a neuron with output,  $y$ , and  $x_i$  as one of the inputs. We will not concern ourselves with other inputs for the moment. The question is: how do we update  $w_i$ , the weight going into the neuron from the  $i$ th input line? First of all, we have agreed to update weights only when there is an error. If there is no error, there is no wisdom in changing the weights, which seem to be doing well for now.

When  $y \neq d$ , we have two cases:  $y = 1$  and  $d = 0$ , or  $y = 0$  and  $d = 1$ . Consider the first case:  $y = 1$ ,  $d = 0$ . This case means the net input to the neuron falls short of the desired value. The "net" is right now negative, which is why "y" is 0; but *net* must be positive to make  $y$  equal 1. Now, how do we change  $w_i$  so as to make *net* move closer toward 0? That depends on the value of  $x_i$ . We must remember that  $x_i$

can take only two values—0 or 1. If  $x_i$  is 0, then it is not contributing to  $y$ , since  $w_i$  and  $x_i$  appear in the formula for net as

$$\text{Net} = w_1x_1 + w_2x_2 + \cdots + w_nx_n.$$

If  $x_i$  is zero, the term  $w_i x_i$ , vanishes. When  $x_i$  and  $w_i$  do not contribute to the output, and therefore to the error, it does not make much sense to change  $w_i$  based on the error. But if  $x_i = 1$ , and the net falls short of 0, we simply need to increase  $w_i$ , to make *net* edge upwards, which is exactly what the equation above does, when  $x_i = 1$ ,  $d = 1$ , and  $y = 0$ .

$$\begin{aligned} w_i &\leftarrow w_i + \eta(1 - 0)1 \\ \text{or, } w_i &\leftarrow w_i + \eta \quad (\text{increase } w_i). \end{aligned}$$

Similarly, when  $x_i = 1$ ,  $d = 0$  and  $y = 1$ , it means that *net* is higher than what it should be. As far as  $w_i$  is concerned, it is pushing *net*, and therefore  $y$ , to undesirably high levels; therefore  $w_i$  has to be reduced from its current value. Again this is exactly what is done by Eqs. (4.3a)–(4.3c) when  $x_i = 1$ ,  $d = 0$ , and  $y = 1$ .

$$\begin{aligned} w_i &\leftarrow w_i + \eta(0 - 1)1. \\ w_i &\leftarrow w_i - \eta \quad (\text{decrease } w_i). \end{aligned}$$

Thus by repeated application of the above three rules of weight modification, for various patterns of Table 4.2, the perceptron learns to respond right to all the patterns and behave like an OR gate.

Naturally, the same learning algorithm can be used to learn other fundamental logic gates also—AND and NOT gates. But the inventor of this elegant learning algorithm, Frank Rosenblatt, did not stop with training these networks on simple logic gates. He trained the networks to learn visual patterns—alphabets, numbers, simple diagrams, and so on. The network could not only learn patterns but could also generalize to slightly new patterns. For example, if it were trained to recognize the alphabet “A” with sloping sides, it could extend its knowledge to understand an A nearly straight sides that arch smoothly at the top. With slight adaptations, the network could also learn to robustly recognize patterns irrespective of slight changes in their position in the input panel, just as the human visual system which would not be fooled by a slight change in position of a familiar pattern, a property of the visual system known as “translation invariance.” These properties of perceptrons quickly made them quite popular.

The secret of the immense popularity of these networks is the fact that they can learn, a capacity that is a unique property of humans and some animals, of biological systems which possess real nervous systems and not of machines. At least not until perceptrons entered the scene. For the first time, we have here artificial systems, based on rather elementary mathematics, but gifted with a power that no machine possessed in the past—the power to learn on its own, the power to perceive, to learn

from examples, and generalize the knowledge to new situations. Very soon a certain aura began to surround not just perceptrons, but neural network models in general.

Rosenblatt's idiosyncratic approach to propagating this new science of learning networks did not help to demystify these models. In science, it is always desirable to present the truth in simple, straight, and precise language—and that precision is often obtained by use of rigorous mathematical language—and state the facts without exaggeration. But Rosenblatt began to make tall claims about the learning capabilities of perceptrons. For example, he would extend the meager lessons learnt from observations of perceptrons to the real brain, treating it as one giant perceptron with a hundred billion McCulloch–Pitts neurons. Basing himself on inadequate knowledge of the neurobiology of the '60s, he claimed that the human brain can store a photographic memory of every image that the eyes can see, at 16 frames/s, for 200 years. As an unfortunate effect of these arbitrary claims, these wonderful learning networks, which had a great potential to revolutionize the study of the brain at that time, actually came to some discredit and harsh criticism.

Two of the strongest critics of perceptrons were Marvin Minsky and Seymour Papert, who ironically are authors of an excellent book named *Perceptrons*. The authors of this book are also pioneers in the exciting field of Artificial Intelligence (AI), that was born just about the time when perceptrons were born. More precisely, the field of AI was born in 1956 at the remarkable Dartmouth Conference more formally known as the Dartmouth Summer Research Conference on Artificial Intelligence. It was organized by John McCarthy and formally supported by Marvin Minsky and other pioneers of AI. The conference is basically a month-long brainstorming session to answer a simple question: "is it possible to make machines that are intelligent?" The conference was based on the conviction "that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." John McCarthy, who coined the term AI in the same conference defines it as: "the science and engineering of making intelligent machines."

The early work of the pioneers of AI, John McCarthy, Marvin Minsky, Alan Newell, and Herbert Simon stood as a strong testimonial of their fundamental philosophy: that it is possible to reduce intelligence, or intelligent behavior to a set of rules. The greater the intelligence is, the more complex the rules. The trick lies in divining the underlying rules of what seems at the first glance intimidatingly intelligent. The AI pioneers and their students wrote astonishing programs that could solve word problems in algebra, or answer logical questions or speak coherent and impeccable English. Less than a decade after the inception of AI, the Department of Defence in the US started funding AI research on a grand scale. AI research labs mushroomed all over the world. Thus at the heyday of AI, perceptrons were just struggling to stake a claim for fame.

In the '60s, AI and neural networks seemed to represent two very different, rivaling approaches to the common question of "what is intelligence?" Whereas one of these approaches tries to reduce intelligence to rules, which must be figured out by a real intelligent person, in flesh and bone, the neural networks exhibit an important characteristic of intelligence, namely, the ability to learn. But the parity ends there.

By mid-60s, AI had already grown to enormous proportions, with a strong financial backing, and practitioners in many countries, while perceptrons are a dream, bordering on a fairy tale, of a single individual—Frank Rosenblatt. The fantastic claims of Rosenblatt about his invention made the matter only worse.

Although Minsky and Papert's criticism of perceptrons must be seen in the social and academic background of those times, their objections to perceptrons, unfortunately, were well founded in mathematics. Perceptrons were not as omniscient as their inventor would believe them to be; they had deep limitations that would seriously jeopardize their acclaimed position as models of brain function and intelligence. Although perceptrons can learn the three fundamental logic gates (and a few other interesting visual patterns), they cannot learn, as Minsky and Papert showed, a slightly more complex logic function known as the XOR.

XOR is a logic function that is a variation of the familiar OR function, and may be stated as “mutually exclusive” or “one or the other but not both.” XOR gate outputs a 1 when only one of the inputs is 1, but not when both inputs are 1. Its truth table is therefore given as

Note that the above truth table differs from that of the OR gate (Table 4.1) only in the last row. The function looks disarmingly simple considering its family resemblance to the OR gate. But Minsky and Papert have shown that perceptrons cannot learn even this simple logic gate. To understand their allegation, we must take a closer look at the exact manner in which perceptrons solve problems.

Let us return to our simple OR gate. We have seen that a perceptron with two input neurons expressed as

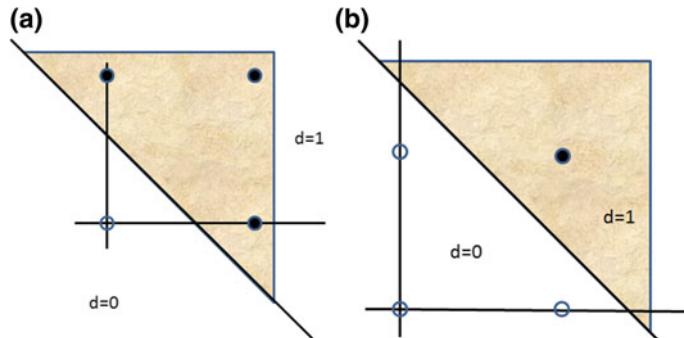
$$y = \sigma(w_1x_1 + w_2x_2 - b)$$

with  $w_1 = w_2 = 1$  and  $b = 0.5$ , behaves like an OR gate. We have verified this mathematically. But we will now verify this geometrically with pictures. Since  $\sigma()$  is a step function, its output,  $y$ , can only take two values—0 or 1. All along we have been allowing  $x_1$  and  $x_2$  also to take only binary values (0 or 1). But there is nothing that prevents us to let them take any real value. Therefore, we can represent the pair of values  $(x_1, x_2)$  as a point on a plane. A binary value,  $y$ , may be associated with every such point by a simple color coding scheme:  $y = 1$  at a point, paint the point white, else paint it black. Now the entire plane is divided into two regions—black and white. We now need to find out the shape of these regions and that of the border separating the two.

Let us recall the definition of the step function

$$\begin{aligned}\sigma(u) &= 1, & \text{for } u > 0, \\ &= 0, & \text{otherwise.}\end{aligned}$$

When  $u$  is positive,  $\sigma(u) = 1$  and when  $u$  is negative  $\sigma(u) = 0$ . The border between the positive and negative regions occurs at  $u = 0$ . Now if we substitute  $u$  with  $(w_1x_1 + w_2x_2 - b)$ , the border occurs when  $w_1x_1 + w_2x_2 - b = 0$ , which is the equation of



**Fig. 4.6** A geometric depiction of how a perceptron models, **a** OR gate, and **b** AND gate

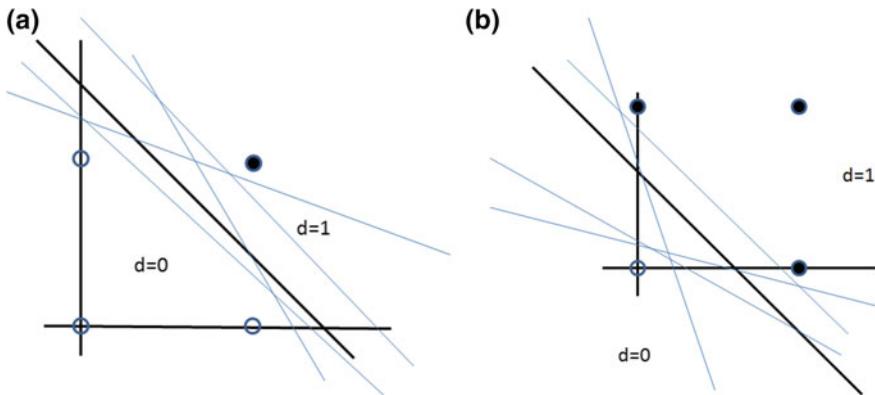
a straight line. Therefore, the black and white regions that correspond to  $\sigma(u) = 1$  and  $\sigma(u) = 0$ , respectively, are separated by a straight line. The slope of the line or its distance from the origin is determined by the neuron parameters— $w_1$ ,  $w_2$ , and  $b$ . The region on one side of this line is white, and the other side is black.

We can now depict the familiar AND and OR gates, as represented by the McCulloch and Pitts neuron, using these shaded and white images (Fig. 4.6a, b). Note that the position of the border (but not its orientation) is the only difference between Fig. 4.6a, b. For the AND gate, the border is shifted further away from the origin. In the AND gate image (Fig. 4.6a), there are three points in the white region, whereas in the OR gate image there is only 1.

Note also that there is no single, unique border that  $y = 1$  points from  $y = 0$  points for either of these gates. For example, there is a whole range of borders that will work for the AND gate (Fig. 4.7a). That is, though we have chosen a particular set of parameters ( $w_1 = w_2 = 1$  and  $b = 1.5$ ) for the neuron to represent an AND gate, there is a whole range of parameters that will work. Similar things can be said about the OR gate too (Fig. 4.7b).

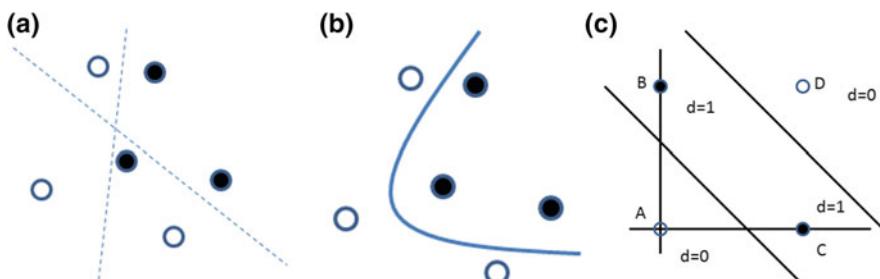
These images (Figs. 4.6 and 4.7) offer an insight into what a perceptron does when it learns a function. Pictorially speaking, a perceptron basically separates two sets of points—corresponding to  $y = 1$  and  $y = 0$ , respectively—with a straight line. Often there is no unique line that can do the separation; there is a whole family of lines. When a perceptron learns a function like, say the AND gate, it might discover any one of the borders, and not necessarily the one we chose: ( $w_1 = w_2 = 1$  and  $b = 1.5$ ). In this sense, the behavior of a perceptron can be described as creative, spontaneous, and not rule-driven, unpredictable since there is no telling which border will be chosen, and almost intelligent. These virtues are shared by other, more sophisticated, neural network models too, as we will see soon.

But then when all is so enchanting and glorious about the perceptron what were Minsky and Papert griping about? There is indeed a deep limitation which can be simply explained using our familiar black and white “decision regions.” When we said that a perceptron separates two sets of points using a straight line, we are making



**Fig. 4.7** A schematic to depict that there are many ways to model, **a** the AND gate and **b** the OR gate, using a perceptron

an important assumption: that it is indeed possible to separate the points in such a manner. We can easily visualize pairs of sets of points that cannot be separated by a straight line. What happens when we demand the perceptron to learn problems that involve such sets of points? The learning process will never end, and the weights of the network will fluctuate endlessly without ever settling down. The network will fail to find a solution because there is none. Even in the cases where the sets of points cannot be separated by a straight line, it is possible to separate them with a *curve* (Fig. 4.8b). But that merit does not apply to a perceptron since the borders in its decision regions are always straight lines. This is exactly the limitation that was exposed by Minsky and Papert when they showed that a perceptron cannot even learn a simple XOR problem. Figure 4.8c shows the four points in the XOR truth table (Table 4.3). It is obviously impossible to separate points A and B versus points C and D using a straight line (Fig. 4.8c). A perceptron will not be able to learn all the patterns of Table 4.3; it usually learns three, but errs on the fourth one.



**Fig. 4.8** **a** An example of two sets of points (filled circles and open circles) that cannot be separated by a straight line. **b** The same pair of sets can be separated by a curve. **c** Data points from the XOR problem cannot be separated by a single line. But they can be separated by two lines

**Table 4.3** XOR truth table

| $x_1$ | $x_2$ | $d$ |
|-------|-------|-----|
| 0     | 0     | 0   |
| 0     | 1     | 1   |
| 1     | 0     | 1   |
| 1     | 1     | 0   |

Thus, the perceptron can only learn a problem when the patterns that constitute the training data are linearly separable. Learning fails when the data is not linearly separable. Although we presented our arguments using pictures, in two dimensions, the same principle applies even general multidimensional spaces, which is the case when the perceptron has  $n$  inputs ( $n > 2$ ). In such cases, we do not talk of separating by straight lines, but by hyperplanes, the linear equivalents in multidimensional spaces. Thus, Minsky and Papert proved that the limitation of the perceptrons arises not just in a few special problems but in a large class of real-world problems. Most real-world problems involve separation of sets of points that are not linearly separable. It is possible to conceive of models that can construct nonlinear (curved) borders and solve such problems too, but perceptrons are not such models.

## Multilayer Networks

Considering these difficulties of perceptrons, Minsky and Papert launched a strong criticism of not just perceptrons, but neural networks in general. To quote from their book Perceptrons:

The perceptron has shown itself worthy of study despite (and even because of!) its severe limitations. It has many features to attract attention: its linearity; its intriguing learning theorem; its clear paradigmatic simplicity as a kind of parallel computation. There is no reason to suppose that any of these virtues carry over to the many-layered version. Nevertheless, we consider it to be an important research problem to elucidate (or reject) our intuitive judgment that the extension to multi-layer systems is sterile.

Such strong criticism rang death knells for future development of neural networks. While their analysis of the limitations of perceptrons (“perceptrons can only solve linearly separable problems”) is true, their prognosis of perceptrons (“...extension to multi-layer systems is sterile”) turned out to be completely false and misleading. And Rosenblatt was not there to address these issues since he died in an unfortunate airplane accident. Such a pessimistic view of the possibilities of neural networks caused a tremendous slowdown in neural network research in the ‘70s. It gave more prominence the “symbolic” approach of AI and marginalized the “subsymbolic” approach represented by neural networks. Contrary to what Minsky and Papert predicted, it was later proved by several workers almost simultaneously that multilayered versions of perceptrons, simply called the Multilayered Perceptrons (MLPs), are free from the weaknesses of a perceptron. But such proof and realization came after a

long decade, in the early ‘80s. The negative criticism was not the sole reason for the delay in progress on the MLPs. There is an inherent technical difficulty in dealing with the MLPs, particularly in discovering a suitable learning algorithm.

To understand this difficulty, we must briefly revisit the perceptron learning algorithm. Weight update depends on two quantities: (1) the error,  $\delta = d - y$ , available at the output, and (2) the input,  $x_i$ . Or, the change in weights is proportional to the product of the error,  $\delta$  (the Greek letter “delta”) and input  $x_i$ . Since the weight update is proportional to the error,  $\delta$ , the perceptron learning algorithm is also known as the “delta rule.” Mathematically the learning rule is trivial to understand; we have given some loose arguments a little while ago, though rigorous mathematical proofs that explain why the learning rule works do exist.

But there is something elegant about this learning rule when seen with the eyes of a computer engineer. Imagine you have implemented the perceptron as some sort of a parallel computer, in which each neuron is a tiny processor. The perceptron is then a network of little computers interacting via the weights, which can also be thought to be implemented by tiny processors. Now when we present an input to the input layer, the neurons of the input layer pass on that information,  $x_i$ , to the single output neuron, via the weights. As  $x_i$  pass over the weights, they are multiplied by the respective weights and the products ( $w_i x_i$ ) arrive at the output neuron. The output neuron in turn sums up all these products, passes the sum through a step function, and computes its output,  $y$ . This completes the so-called forward pass. Now we update all the weights by a reverse pass. The first step in the reverse pass is computation of the output error,  $\delta = d - y$ . This single value,  $\delta$ , available at the output neuron, is received by each of the weights,  $w_i$ . Likewise, the inputs  $x_i$  also are received by the corresponding weight  $w_i$ . The weight combines the two quantities received from the output end ( $\delta$ ) and the input end ( $x_i$ ), and computes their product to update itself. This ends the reverse step. Thus, in the reverse step, when the learning occurs, the error from the output layer is propagated backward, (or downward, depending on which way you draw your networks!) toward the input layer. It is noteworthy that all these computations are *local*. Every unit in this whole network, either a neuron or a weight, needs to only interact with the units with which it is physically connected to perform the computations it needs to perform. Each neuron has only to receive inputs from neurons on its input side, and broadcast its output to the neurons connected to it on the output side. Similarly, the weights have only to keep track of the quantities that are available at its two ends—the output end, where the error ( $d - y$ ) is available, and the input end, where the  $x_i$  is available. This locality property comes in handy if you are going to implement the network as a parallel computation, where each neuron and weight is represented by a tiny processor.

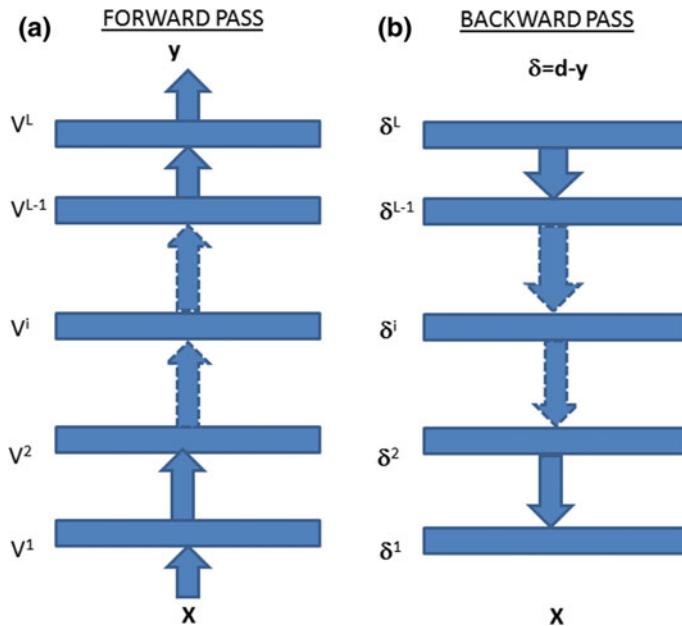
But when we try to reenact the same steps—forward and backward passes—over a network with more than two layers, like an MLP, we run into some serious difficulties. The difficulty arises in designing a learning algorithm that can work with three-layer networks. The weight update rule for the perceptron, which is a two-layer network, updates a given weight using the product of the error,  $d$ , at one end of the weight and the input,  $x_i$ , at the other. The same rule becomes meaningless when applied to three layer networks.

In a three-layer network, in addition, the input and output layers, as in a perceptron, there is also an intermediate layer, known as the *hidden* layer, since it is “hidden” from the view of the inputs and outputs. We can only control the network by presenting inputs, and prescribing desired outputs. We have no direct access to the innards of the network: its weights. Our inputs and desired outputs indirectly affect the weights and teach the network a suitable task. This arrangement is meant to simulate how the brain learns. If we want to teach ourselves something new, like, say, a foreign language, we do not have the option of directly inscribing the words and sounds of the foreign tongue onto the synapses of Wernicke’s and Broca’s areas and prepare rapidly for an upcoming tour! By a slow and measured exposure to language guides and tapes (input/outputs), we hope to indirectly affect our brain (the massive “hidden” layer).

Now we may note that the error,  $\delta$ , is well defined for the output layer, but not for the hidden layer since there is no such thing as a desired output for the hidden layer. The responses of the neurons of the hidden layer can be anything: the only constraint is on the output layer’s response, which should resemble the desired output,  $d$ . Since no suitable  $\delta$  can be defined for the hidden layer, it is not clear how to train the weights of the first stage of weights from input layer to hidden layer. This is the so-called credit assignment problem termed insurmountable by Minsky and Papert.

Notwithstanding the debate about who should take the blame, the ‘70s brought a lull in neural networks research. Thanks to the stigma associated with neural networks, not much has happened in these “quiet years” as some would call them. Even though a few valiant souls worked even in this difficult period, they worked mostly “under the hood” and their work did not receive the attention it deserved until the time is ripe to pop the hood and come out into the open. After a decade-long near inactivity, the ‘80s began with explosive developments in neural networks research. Three different groups/individuals developed the famed backpropagation algorithm, the algorithm for training MLPs, all within a narrow span of a little more than a year. David Parker described the method he developed in a technical report at MIT in 1985. Yann LeCunn published his work in a French conference on Cognition in 1986. The third instance of the discovery was not by an individual, but a group of researchers from the Computer Science Department of Carnegie Mellon University (CMU): David Rumelhart, Geoffrey Hinton, and Robert Williams. At about the same time, in retrospect, it was also discovered that a solitary hero known as Paul Werbos discovered actually a more general version of the backpropagation algorithm back in 1974 and described it in his Ph.D. thesis at MIT in applied mathematics. Paul Werbos now generally enjoys the credit of priority in the discovery of the backpropagation algorithm. But the work of David Rumelhart and his colleagues deserves a special mention since they were responsible for popularizing the method they discovered by applying it to a large variety of problems in psychology and paving the way to a new approach to understanding brain function and cognition, an approach that is termed *Parallel and Distributed Processing (PDP)*.

The breakthrough of the backpropagation algorithm consists in solving the problem of the absence of a suitable error,  $d$ , for the hidden layer. The inventors of this algorithm solved the credit assignment problem by defining a new form of  $\delta$ , that does not require an explicit desired output,  $d$ , but is simply expressed in terms of



**Fig. 4.9** The **a** forward and **b** backward passes of the backpropagation algorithm

the  $\delta$  available at the output layer. Once we have the  $\delta$  for the hidden layer—let's call it  $\delta^h$ —we can express the update rule for the first stage weights, the weights connecting the input layer and the hidden layer, in terms of the product of  $\delta^h$  and  $x_i$ . We now have a rule for updating all the weights of a three-layer MLP.

The same approach applies to an MLP with any number of hidden layers (Fig. 4.9). In such a network the input  $x$ , is passed on the first hidden layer, where the responses (denoted by  $V^1$ ) of the neurons is computed. These responses ( $V^1$ ) are passed on to the second hidden layer and neural responses of that layer ( $V^2$ ) are computed. This procedure is continued until at the last layer, the network output,  $y$ , is computed. This completes the so-called forward pass (Fig. 4.9a). Once  $y$  is available, using the desired output,  $d$ , we compute  $\delta = (d - y)$  at the output layer as we did before in case of perceptrons. In this case,  $\delta$  at a given hidden layer depends on the  $\delta$  at the hidden layer just above it. This chain of dependencies extends, recursively, all the way up to the output layer, where the original  $\delta$  is computed. It is less confusing to use a more suitable notation to represent  $\delta$ 's at various layers. In a network with  $L$  layers,  $\delta^L$  denotes the error ( $= d - y$ ) at the last or the output layer;  $\delta^{L-1}$  denotes the error at the layer just below the output layer, i.e., the last hidden layer; and finally  $\delta^1$  denotes the error at the first hidden layer (Fig. 4.9b). Dependencies involved in the computation of  $\delta$ 's may be depicted as

$$\delta^L \rightarrow \delta^{L-1} \rightarrow \delta^{L-2} \rightarrow \dots \rightarrow \delta^1.$$

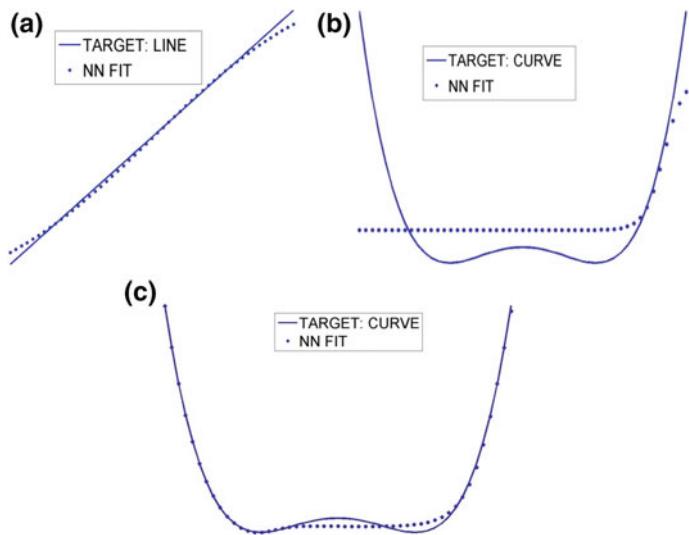
Once the deltas ( $\delta^L, \delta^{L-1}, \dots, \delta^1$ ) are all computed updating the weights is straightforward. A given weight between two layers (say,  $m$  and  $m - 1$ ) is updated using the product of the delta available at  $m$ th layer ( $\delta^m$ ) and the neuron response  $V^m$  available at  $(m - 1)$ th layer. In summary, the backpropagation algorithm involves a forward pass of the input over the layers all the way to the output layer, where  $y$  is computed and a backward pass of deltas or errors from the output layer all the way to the first hidden layer, in order to compute the deltas at every hidden layer. This backward propagation of errors or deltas justifies the name backpropagation algorithm.

The existence of a learning algorithm for MLPs is not an adequate cause for celebration. The real motivation for the study of MLPs is the possibility of overcoming the weaknesses of perceptrons, which can only classify linearly separable problems. The existence of a learning algorithm does not guarantee better performance. Just as Minsky and Papert performed a thorough mathematical analysis of perceptrons and demonstrated their deficiencies, someone had to study MLPs mathematically and investigate their strengths.

One of the first results of that kind came from George Cybenko a mathematician at Dartmouth College. His result is a theorem, known as the Universal Approximation Theorem (of MLPs) which states that an MLP can learn problems of arbitrary complexity provided that it has a hidden layer with an arbitrarily large number of neurons. Thus, the theorem links the complexity of the problem at hand to the size of the network required (specifically the size of the hidden layer) to learn that problem up to the desired accuracy. For example, a three-layer MLP required to classify the 26 alphabets of English is likely to have more neurons in its hidden layer than a network that is required to classify the 10 digits (0–9). A network that must classify the 50+ main characters (vowels and consonants) of a typical Indian language script, like Devanagari, for example, with its considerably more complex and ornate characters, is likely to require yet larger hidden layers.

Figure 4.10 shows the result of a simple exercise that can be performed on a computer using a standard neural network software. A three-layer MLP is used to learn two functions—(1)  $y = x$ , and (2)  $y = x(x - 1)(x - 2)(x - 3)$ . The functions are considered only in the following range of values of  $x$ :  $[-1, 4]$ . Observe that an MLP with just 1 hidden neuron in the hidden layer is sufficient to learn the simple linear function ( $y = x$ ) (Fig. 4.10a). The same sized network when applied to the second function (nonlinear, with four roots) obviously gives poor results (Fig. 4.10b). But when the number of hidden neurons is raised to 4, the network is able to learn the second function too. The network's output is now able to follow the winding contours of the second function satisfactorily. This link between the complexity of the problem and network size provides a powerful insight since it suggests an interesting significance to brain size, an issue we tried to grapple with in Chap. 2.

Armed with a powerful learning algorithm with which networks of arbitrary size can be trained, and emboldened by a powerful “Universal approximation theorem” that guarantees the existence of a solution for arbitrarily complex problems, a whole vanguard of explorers began to apply MLPs to a variety of situations and probe the limits of their capabilities. These networks were trained to read alphabets, handwritten digits, read aloud English text, articulate sign language, recognize sonar signals,



**Fig. 4.10** **a** A MLP with a single neuron in the hidden layer learns to fit a straight line (solid—target line; dashed—MLP output). **b** The same network is trained to fit a polynomial with poor results (solid—target curve; dashed—MLP output). **c** But the network with 4 hidden neurons is able to learn the polynomial fairly well

learn English grammar, or drive a car. The list is quite long. Every application revealed something new and interesting about the way these network models learn to solve problems. The applications also indicated how the human brain might be solving the same problems.

Most noteworthy among the pioneering attempts to apply MLPs (and other neural networks that were subsequently developed), are those that were pioneered by a group of researchers, who called themselves the Parallel and Distributed Processing (PDP) group. The PDP group consisted of a motley group of researchers with a variety of backgrounds—physics, neuroscience, computer science, mathematics to name a few. But all of them had a shared interest—to study the brain as a parallel and distributed system, and express its computations as an outcome of interactions among a large number of small units called neurons. The idea of regarding the brain as a parallel and distributed processing system is not new. Back in the nineteenth century, summarizing the lessons learnt from various aphasias, Carl Wernicke described brain exactly as a PDP system and tried to resolve the local versus global debate through such a synthesis. But it was only a conceptual synthesis, which was what was possible in an era that was nearly a century before the advent of computers. But comparatively, the late '60s and '70s offered greater opportunities, though quite rudimentary by current standards, to study network models of perception, cognition, and action through computer simulations. The earliest seeds of PDP work were sown in 1968, at the University of California at San Diego, where four inquisitive young researchers—Geoffrey Hinton, James Anderson, David Rumelhart, and

James McClelland—began to study, using network models, how people perceive words. Over more than a decade, this small initial group worked on a variety of problems related to cognitive science within the PDP framework. In the ‘80s, the PDP group grew beyond the boundaries of UCSD, with kindred individuals working on PDP lines at Carnegie Mellon University, MIT, and other institutions. In 1986, some of the key ideas and results of PDP group, developed over more than a decade, were published as two volumes titled: “Parallel Distributed Processing: Explorations in the Microstructure of Cognition.”

The word “microstructure” that appears in the caption deserves special attention. It exemplifies the distinct approach to cognition taken by the PDP group, in contrast to the more classical AI approach, according to which all cognition can be reduced to a set of rules, that lend themselves to an explicit, unambiguous definition. The challenge lies in discovering those rules for a given cognitive task—but there are always rules. In the PDP approach to cognition, there are no rules; the function of interest is encoded in the interactions among the neurons. It is coded implicitly in the connection strengths, and therefore it may not be always possible to make explicit the knowledge, the meaning, “the microstructure of cognition” associated with a connection strength.

## Learning Past Tense

It is easiest to illustrate the contrast between the rule-based approach and the neural network approach to cognition as was championed by the PDP group, with an example. Rumelhart and McClelland studied how children learn past tense and asked if similar behavior can be reproduced by neural network models. Learning past tense of English verbs is a challenging problem since there is no general rule: the standard “d” or “ed” ending does not work for all verbs. Thus verbs whose past-tense forms have the standard “d” or “ed” ending are known as the *regular* verbs, while the exceptions (e.g., come and came) are the *irregular* verbs. Linguists who considered this problem have identified three stages in the manner in which children learn past tense of verbs. In the first stage, children use a small number of verbs in the past tense. Most of them are irregulars, but tend to be high-frequency words, like, for example, come (came), get (got), give (gave), look (looked), need (needed), take (took), and go (went). Two of these seven verbs are regular while the rest are irregular. At this stage, it appears that children know a short list of disparate items, without reference to any underlying rule.

In the second stage, children begin to demonstrate implicit knowledge of linguistic rules. They begin to acquire a large number of verbs in this stage, with a good percentage of them being regulars (e.g., answer and answered). It has been shown that in this stage the children simply do not memorize past-tense forms since they tend to supply the standard “d” or “ed” ending to new verb forms. For example, it was observed if the children are encouraged to use a fabricated verb like *rick*, and try to invent its past-tense form, they naturally arrive at *ricked*. But the trouble, in

this stage, is that children tend to over-generalize and wrongly supply the standard ending to even irregular verbs learnt well in the previous stage. For example, they might start using *comed* or *camed* as past-tense forms of *come*.

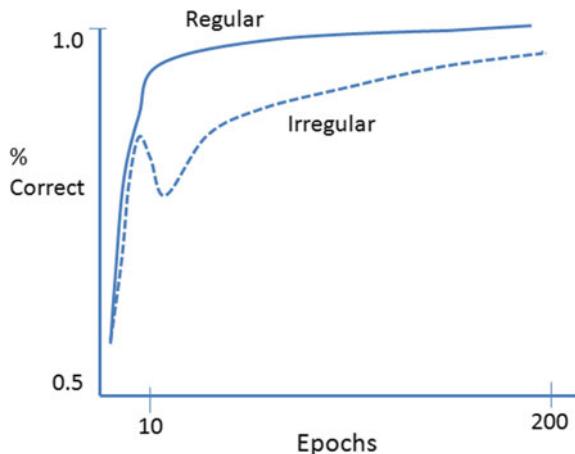
A resolution of this conflict seems to occur in the third stage in which both regular and irregular forms coexist. In this stage, the child's vocabulary grows further, with the addition of a larger stock of both regulars and irregulars. But they somehow learn to use the right ending for regular verbs and avoid the wrong generalization in case of irregulars. The irregular verbs which were associated with the wrong past tense in the second stage are now corrected.

Rumelhart and McClelland decided to study these trends in past-tense learning using an MLP. The network is trained to accept a verb form in present tense as input, and produce the past-tense form as output. In order to imitate the stages by which children learn, they trained the network also in three stages. In each stage, the choice and number of verbs used for training reflect the stock of verbs to which children are exposed to in that stage. In the first stage, the network is trained on the following 10 high-frequency verbs: *come*, *get*, *give*, *look*, *take*, *go*, *have*, *live*, *feel*, and *need*. Of these, eight are irregulars and only two are regulars (*live* and *need*). The network was trained for 10 epochs on these 10 verbs, which were learnt successfully. This stage of learning is compared to Stage I of past-tense acquisition in children. Now, to simulate Stage II, 410 additional medium-frequency verbs were added to the original shortlist of 10. Among the 410 verbs, 334 were regular and 76 were irregular. The network was trained for 190 epochs on this expanded word list of 410 verbs. (1 epoch = 1 complete presentation of all the patterns in the training set). This course of training is said to be analogous to Stage II. After this stage, the network is not trained any further and the weights are frozen. The network is only tested on a new stock of 86 low-frequency words consisting of 72 regulars and 14 irregulars.

Figure 4.11 shows the results of network training, with two separate graphs depicting performance over regular and irregular verbs. Performance over the stages of training, I and II, with 200 (=10 + 190) epochs is considered. Over the first 10 trials where the network only learns a small number of regular and irregular verbs, there is hardly any difference between the two curves. However, at the beginning of stage II, when 410 new verbs are introduced, note that performance over irregulars showed a sudden dip, while that of regular verbs showed a continued increase. On further training, performance over irregular verbs showed a steady increase, approaching but matching the performance of regulars throughout training. Therefore, the performance over irregulars, relative to the regulars, shows a drop at the beginning of stage-II, picking up again toward the end of the stage, resembling the familiar "U-shaped learning curve" that is generally associated with learning of the irregular past tense.

An important feature of the past-tense learning network is its deviation from the rule-based approaches that were rampant in the days before neural networks came in existence. The past-tense network was able to learn the "rule" of the standard "d" or "ed" ending of regular verbs and also reconcile that rule with a large number of "exceptions." The rule-like behavior of the trained network is only an appearance: the network learnt the regularities ("rules") and the less frequent cases ("exceptions")

**Fig. 4.11** Performance of the network training on past-tense learning. Solid curve represents performance over regular verbs, while the dashed curve denotes performance over irregulars



present in the training set. But basically, the network was tested if it can learn a single rule and distinguish it from a host of exceptions. Drawing inspiration from this pioneering study, a bold attempt was made to apply MLP to a much harder problem which involves a large number of both rules and exceptions.

## NETtalk: A Network that Can Read

Reading the English language poses quite significant challenges to a young Indian reader, because English, unlike Indian languages, is not a phonetic language. English is not a case of “what you see is what you read.” The sound associated with a character depends on its context. Consider, for example, the variety of pronunciations that the letter “c” enjoys in the following words: chip, chic, coal, and cerebrum. Or, mark the variations in the pronunciation of the first vowel “a” in: gave, halt, toad, paw, and cat. Soon after the development of the backpropagation algorithm in 1986, Sejnowski and Rosenberg, trained an MLP to read aloud English text, a first shot at using neural networks for such a task. Perhaps the only other contender for fame, at that time, was a rule-based system known as DECTalk. This system had a large database of words with pre-programmed pronunciations. When an unfamiliar word is encountered, the system used a set of rules for converting the characters into phonemes, the smallest set of identifiable sounds into which speech can be segmented. Whenever DECTalk encountered new words on which its rule set failed to produce correct pronunciation, a human expert had to intervene to add new words, and new rules for converting text to phonemes.

The system designed by Sejnowski and Rosenberg, called the NETtalk, consisted of an MLP with three layers: a short text segment is presented to the input layer, and the phoneme recognized is read off the output layer containing 26 neurons. The

hidden layer consisted of 80 neurons. Since pronunciation of a letter depends on its context, letters are presented along with a brief context (three letters before and three after). Therefore, strings of 7 ( $3 + 1 + 3$ ) are presented to the input layer of the network. The 26 letters of the English alphabet along with three punctuation marks (comma, full stop, and a “blank space” that separates words) are represented in the input. The phonemes used to train the network were sliced from a corpus of informal, continuous speech of a child. A set of 1000 words were chosen from a large corpus of 20,012 from a dictionary. This set of words were presented repeatedly for several epochs<sup>1</sup> until the network showed sufficiently high accuracy on the training set.

Many valuable lessons were learnt by the researchers as the network learnt to read aloud the training words. A good number of words were pronounced intelligibly after only a few epochs, and after about 10 epochs the text was understandable. One of the first signs of learning exhibited by the network is the distinction between vowels and consonants. But all vowels were pronounced as a single vowel and all consonants were read out as a single consonant. The resulting speech sounded like the babbling of infants (baba, mama,...). The next development observed was recognition of word boundaries: the network paused briefly at the space between words, which gave the sense that speech consists of discrete units corresponding to words, and is not an inchoate mass of continuous noise. Even the errors committed by the network were natural, and almost human-like, and not totally random. For example, a consonant like /th/ as in “thesis” sounded like a similar sound that occurs in “these.”

An important, brain-like characteristic exhibited by the NETtalk system is a robust performance in face of damage. Brains are known for their legendary ability to repair and reorganize themselves post-injury so that intact parts take over the function of the damaged portions. Emergency units witness every day how brains recover almost miraculously from damage, from minor contusions to traumatic brain injuries, so as to recover the function that would have been lost permanently. The classic case of Phineas Gage, a story often recounted in histories of neuroscience, is an excellent testimonial. Gage was a railroad construction worker, who lived in the late nineteenth century. While at work, he met with a fatal accident in which a crowbar flew in his face and shot through the frontal area of his brain. He survived the accident, and, though deemed unfit to continue with his professional work, he led an otherwise normal life.

Brains are capable of exhibiting such extreme robustness against damage simply because of their massive network structure, and their ability to relearn and rewire. A hint of this ability is also seen in the NETtalk system. Random noise was added to the weights of the trained NETtalk system. Up to a noise level of  $\pm 0.5$ , there was practically no change in the network’s performance. As noise amplitude was increased, performance also degraded gradually. This property, known as graceful degradation, is shared by neural network models in general.

Another related strength exhibited by NETtalk is the rapidity with which it recovered from the damage. When the random noise of amplitude greater than 0.5 was added to the network’s weights, the network showed a sudden reduction in perfor-

---

<sup>1</sup>1 epoch = a presentation of all patterns in the training data set.

mance. When the network was retrained from that point, performance increased much more rapidly than it did the first time it was trained.

Thus, it was possible to capture several brain-like virtues using small networks consisting of just three layers, and a few hundred neurons (trivially smaller than the brain which has a billion times more), thanks to the elegant learning algorithm, and the highly “neural” qualities of distributed representation and parallel processing of information. These simple network models prepared the ground for the creation of more complex, elaborate models that gradually approach real nervous systems in complexity.

There are many different ways in which one can visualize the creation of more elaborate and realistic models. One can, for instance, consider networks with more layers than three, and study more realistically several important systems in the brain. Even the simple reflex action, instantiated in the sudden, explosive response we produce when we step on something sharp, a response that is considered quite simple in nature, involving only very local and limited parts of the nervous system, actually involves several stages of neural processing. And if we consider how we see and recognize objects in the real world, we need to describe systems with more than 10 stages of neurons, a number that can increase even further, depending on how we count stages. But adding more stages is perhaps not the best way to make progress on our journey to demystify the brain. It is still more of the same kind. We must consider networks that are different in kind.

The networks introduced in this chapter—perceptrons and their multilayer versions—are all of one kind. Information on these networks always propagates in a single direction—forward. There is a unidirectional flow of information from the input to the output layer, through a variable number of hidden layers. But real brains are full of loops. When I look at a cup and pick it up, visual information originating in my retinae does not March, in a single rank and file, undergoing rigidly sequential transformations, leading up to the activation of muscles of my hand. Information circulates, coursing through loops, great and small, before accumulating in the appropriate motor cortical areas, and ultimately driving the muscles of my hand into action. Networks with loops are radically different from the non-looping, or feedforward networks like the MLPs. They bring in new effects, explains very different phenomena. For starters, they explain how brains store memories, one of the most important functions we use our brains for. Such networks with feedback loops motivate the discussions of the following chapter.

## References

- Buchanan, B. G. (2005, Winter) A (Very) brief history of artificial intelligence. *AI Magazine*, pp. 53–60. <http://www.aaai.org/AITopics/assets/PDF/AIMag26-04-016.pdf>.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function mathematics of control. *Signals, and Systems*, 2(4), 303–314.
- Kumar, S. (2004). *Neural networks: A classroom approach* (736 pages). New York: Tata McGraw-Hill Education.

- McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 7, 115–133.
- Minsky, M. L., & Papert, S. A. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. Washington, DC: Spartan Books.
- Rumelhart, D. E., & McClelland, J. L. (Eds.). (1982). *Parallel distributed processing: Explorations in microcognition* (Vols. 1 and II). Cambridge, MA: MIT Press.
- Sejnowski, T. J., & Rosenberg, C. R. (1988). NETtalk: A parallel network that learns to read aloud. In J. A. Anderson & E. Rosenfeld (Eds.), *Neurocomputing foundations of research* (pp. 663–672). Cambridge, MA: The MIT Press.

# Chapter 5

## Memories and Holograms



*Memory, inseparable in practice from perception, imports past into the present, contracts into a single intuition many moments of duration, and thus by a twofold operation compels us, de facto, to perceive matter in ourselves, whereas we, de jure, perceive matter within matter.*

—Henri Bergson, in Matter and Memory.

### Shocks that Elicit Memories

A woman in her 40's lay on the surgical table of neurosurgical ward in the Christian Medical College (CMC), Vellore, a small, unremarkable town in the southern part of India. The woman, most probably from one of the northeastern states had traveled a long way to Vellore, obviously to access the superior medical services offered by CMC. A member of the team of surgeons that surrounded the woman was asking her to count numbers from 1 through 10. She began to count aloud with her strong northeastern accent but stopped suddenly midway as though she was interrupted by an invisible force. That force was the electrical stimulation—mild shocks—delivered by another member of the team to Broca's area, a brain region responsible for control of our speech, typically located in the left hemisphere in right-handed people. Shocks delivered to this area by the surgeon interfered with ongoing counting. The surgeon placed a tiny piece of paper with a number printed on it, at the spot where they found a moment ago where stimulation stopped speech. The piece of paper is the surgeon's landmark for Broca's area.

The very sight of a person talking, or for that matter doing anything normally, with a part of the scalp pinned to an external support, with a piece of the cranium sawed and taken out, and half of the brain exposed, might seem surreal to the uninitiated, but not an uncommon scene in neurosurgery operation theaters all over the world. There are occasions when neurosurgeons choose to operate on the brain while the patient

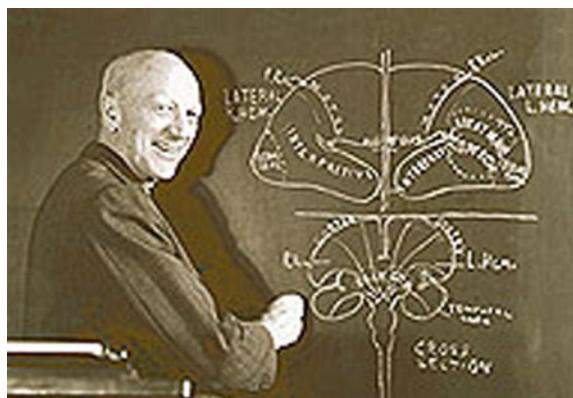
remains conscious, “cooperating” with the surgeon, meekly doing his/her bidding, be it counting digits or twiddling fingers, not out of a morbid intent to impose unearthly torture on an already suffering individual. The choice is made out of a simple surgical exigency. While marking the brain tissue that has to be removed, the surgeon has to ascertain that intact tissue is maximally spared. Unfortunately, often the affected tissue appears quite similar to intact tissue, leaving only one option to the surgeon: to distinguish between normalcy and dysfunction by electrical stimulation. When the probing electrode of the surgeon hit a spot that was able to halt the ongoing counting, the surgeon discovered an intact Broca’s area.

Surgeries of this kind, which experts fondly call “brain mapping by electrical stimulation,” were pioneered by Wilder Penfield, a Canadian-American neurosurgeon who worked in the earlier part of the last century (Fig. 5.1). Penfield sought to find a surgical cure to epilepsy, a devastating disease that resisted pharmacological treatment. Epilepsy is a condition in which electrical signals in the brain go awry. The regulated background chatter in a normal brain is displaced by large synchronized waves of neural forest fire that often originates from certain vulnerable spots—the “epileptic foci”—and spread to brain areas far and near. These bouts of uncontrolled neural activity, termed seizures, may last only a few seconds or, in extreme cases, may continue unabated for hours or days until subdued by strong pharmacological intervention. A mild seizure may be experienced as a brief spell of unconsciousness or it may precipitate into a full-blown convulsion.

Penfield explored surgical options to control intractable epilepsy. His strategy was to locate epileptic foci, through systematic electrical stimulation, and lesion the focal area, thereby cutting off the problem at its roots. These explorations led to him mapping various regions of brain’s surface, particularly temporal lobe where epileptic foci are often located. Very often patients reported vivid recall of past auditory experiences, most probably because superior temporal lobe is the site of auditory area, a part of the brain that processes sounds. For example, the patient might hear the voice of his cousin in Africa, or recall the well-known march of *Aida*. The experience of recall is so vivid and living that the sounds seemed to originate from the immediate vicinity of the patient. Yet the patients were not overwhelmed by the hallucination and were aware of their real, immediate surroundings—the operating room, the surgeon, and the intimidating instrumentation. It was as though the patients had a double stream of consciousness at the time of stimulation, one corresponding to their real surroundings, and the other pertaining to the hallucinatory auditory experience.

Similarly, when borderlands of parietal and temporal lobes were stimulated, the patients reported complex visual experiences. More of such spots were discovered in the right hemisphere than in the left hemisphere. When these spots were stimulated, patients re-experienced vivid scenes from their past. On subsequent enquiry, the patients were able to clearly link their experiences on the surgical table, with actual past events in their life (Fig. 5.1).

These findings led Penfield to believe that memories of one’s past are stored, as in a tape recorder, in specific brain sites. It is when these sites are electrically stimulated that the patients recall the corresponding memories. Thus, Penfield’s research

**Fig. 5.1** Wilder Penfield

supported a “localized” view of organization of memory. Memories are stored at specific sites in the brain and can be recalled by stimulating those sites.

But quite a contrary view on the subject of memories was arrived at by Karl Lashley who sought to find the sites of memories, by performing on rats, experiments that are far bolder than those performed by Penfield on humans (Fig. 5.2). A big goal of Lashley’s research was to search for the *engram*, a word that denotes the biochemical and biophysical machinery for storing memories in the brain. Penfield’s work with humans revealed that brain stimulation at specific sites activated these engrams and retrieved memories. Lashley sought to accomplish the same in experimental animals. Early in his career Lashley came under the influence of John Watson, a pioneer in behaviorism. At a time, when the techniques of brain research were not sophisticated enough to provide detailed structural and functional information about the brain, neuroscientists attempted to fill the vacuum by subjective, speculative accounts of how the brain performs its functions. Behaviorism emerged as a reaction to this unhealthy trend and insisted on casting one’s theories of mind/brain strictly in terms of observables and quantifiable aspects of behavior. For example, thou shalt not say that an animal pressed a bar vigorously; a behaviorist would say that the animal pressed the bar so many times a second. Trained in such a tradition Lashley set out to study the link between brain damage, which he tried to quantify, and behavior (Fig. 5.2).

Although Lashley worked with a variety of animals, his most eminent work was on maze learning in rats. Rats are known for their remarkable ability for finding their way out of complex mazes. Thanks to this special gift, rats have been used for over a century for studying how the brain represents and navigates through space. Lashley exposed rats’ brains and made long straight cuts through the surface in various regions. The total length of the cuts is treated as a control parameter. As this “parameter” is varied the efficiency with which rats escaped from the maze is noted. What Lashley observed, to his astonishment, is that the escape efficiency depended more or less strictly on the total amount of damage, not much on the location of the damage. This result flies in the face of earlier work of Penfield and others who

**Fig. 5.2** Karl Lashley

concluded that brain function, and memories, are localized. Thus Lashley's search for a precise location of the engram ended in a failure.

Lashley summarized his experimental findings into two "laws" of brain function. The first law, the law of mass action, states that the cerebral cortex works as a whole and not as a patchwork of modules each working independently. The second law, termed the principle of equipotentiality, states that if a certain part of the brain is damaged, other parts may reorganize themselves so as to take up the function of the damaged region.

Considering the strong reductionist tendencies of twentieth-century science in general, and the "atomistic" thinking (one gene → one protein → one phenotype or one germ → one disease → one drug) that was strong in biology, it would have been easier to accept localization in brain, than to appreciate a theory of brain function, or of engram, that is delocalized. Paradoxically, localization and delocalization are equally true and important facets of the brain. A single neuron in visual cortex responds only when a bar of a certain length and orientation is placed precisely at a certain spot in the visual space—a case of extreme localization. There have been cases when a whole hemisphere has been removed in childhood, and the individual grew up almost normally, went to college and led a successful life—a case of extreme delocalization.

## Memories as Holograms

One of Lashley's students, Karl Pribram, thought about his mentor's findings seriously and decided to take the idea forward. He started looking for an appropriate model that can explain the distributed, delocalized nature of engrams. One day he chanced upon an issue of *Scientific American* that carried an article on holography and its recent successes. In popular accounts, holograms are often described as "3D photographs" in contrast to the familiar two-dimensional photographs. An ordinary photograph, with only length and breadth dimensions, looks the same whichever

angle you see it from. Contrarily, a real three-dimensional object, with the added depth dimension, reveals hidden features as you move around it seeing it from different angles. Viewing a hologram is similar, in a limited sense, to viewing a real, 3D object. It reveals hidden features as you change your vantage point.

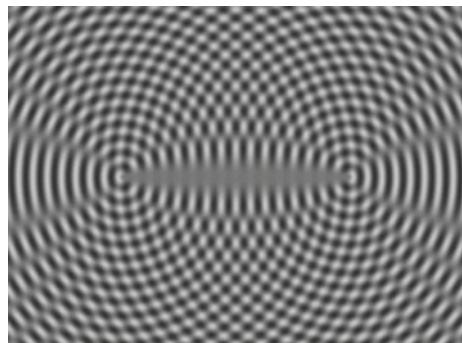
Another property of a hologram is the manner in which it responds to damage. When a 2D photo is damaged locally—say a local burn or an ink spill—the specific part of the image is destroyed forever. Information in a photo is “localized.” But in a hologram, local damage degrades the entire image to a small extent—thus information in a hologram is delocalized. This aspect of a hologram intrigued Pribram since he saw a possible analogy to the engram that he was just searching for. He decided to learn more about holograms from his son, who was a physicist.

Mechanisms that underlie creation of a normal photographic image can be explained using geometric or ray optics, the earliest version of optics which every highschooler learns. In geometric optics, light is described as a bundle of vanishingly thin “light rays.” These rays travel in straight lines until they bounce off reflecting surfaces or bend at the interfaces between unlike media. When we see an object, rays from a light source fall on the object, bounce off the object’s surface, and pass through the pupil of our eye to form an image at the back of the eye, on the retina. If we replace the eye with a camera, we can understand the formation of an image on the photographic film. This “ray” story is adequate to explain the formation of the image on the retina, or on a photographic film, but insufficient to explain the physics of the hologram.

It turns out that the ray description of light is only an approximation valid at sufficiently large length scales. A more accurate description of light is as a wave—not a material wave that we observe on the surface of a lake, but a wave, or a propagating disturbance, in an electromagnetic field. In the world of “ray” optics, the ray coming from an object conveys several properties about the spot on the object from which it originates—intensity (is the spot bright or dim?), direction (which way is the spot located?). In the world of wave optics, the wave not only carries information about intensity and direction but in addition, carries a new property known as phase, which has no counterpart in the “ray” world.

Simple domestic experiments with a large tub of water can reveal a few things about waves. If you stick your finger suddenly in a large tranquil tub of water, you will notice that circular waves originating from the spot where you dunk your finger, and expanding in all directions. If you just looked at a single spot on the surface of the water, you could notice water going up and down. There is no real movement of water, as you can test by dropping a tiny piece of paper, and watch it bob up and down at a single spot. If it were a mere gentle dip you would see that the paper makes only a small oscillation. If you struck harder, making something of a splash, you will notice that the paper swings up and down by a greater extent. This extent of the wave is known as its amplitude, a property that is related to the “energy” of the wave. Another property of the wave is the rate at which a point on the wave bobs up and down—its frequency. In case of a light wave, this frequency refers to the color of light. Red, for example, has a lower frequency than blue. A third property of the wave, the most relevant one for our present purpose, is phase. It refers to the state of

**Fig. 5.3** Interference pattern produced when two sets of circular waves originating from two points meet and interfere



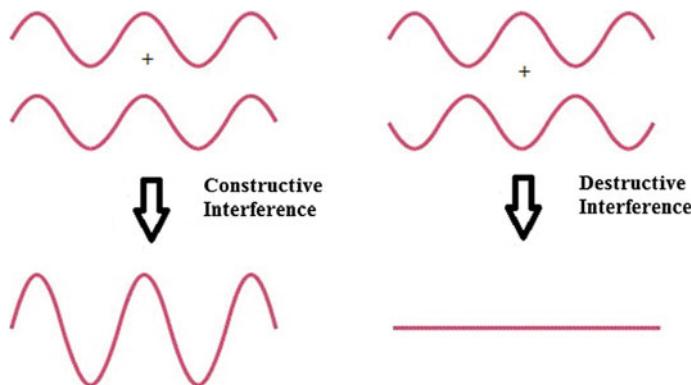
oscillation of a wave in its cycle. By analogy, we may consider a season to represent the phase of a year.

A phase is a property of a single wave. Now it is possible to compare the phases of two waves and talk of phase difference. Imagine two joggers going around a circular track running at equal speeds. Jogger A, however, had a head start over jogger B, and continues to lead B forever since their speeds are the same. Thus, we say that A's phase is greater than that of B, or A's phase leads B's phase. If B had a head start then we say that A's phase lags that of B. Phase difference is the basis of another important wave-related phenomenon known as interference, a phenomenon that is closely related to holograms.

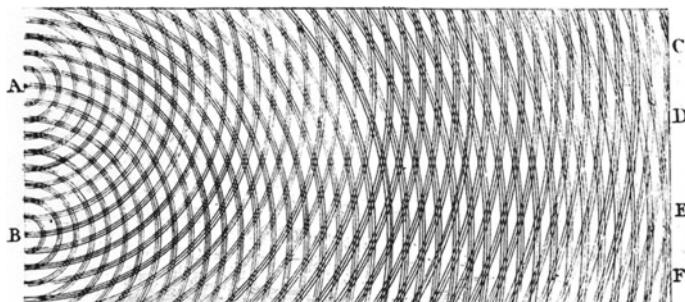
It is straightforward to observe interference in our familiar tub of water. Instead of sticking a single finger, stick two at different ends of the tub, simultaneously. You will watch two sets of expanding, circular waves originating from two different spots. These expanding waves meet somewhere in the middle of the tub, like clashing armies, and form interesting grid-like patterns of troughs and crests on the water surface. This meeting of waves is known as interference, and the patterns such trysting waves form are called interference patterns (Fig. 5.3).

If the two waves meet such that the crests of either wave coincide (or the phase difference is zero), then the two waves add up, leading to what is known as constructive interference (Fig. 5.4a). When the two waves meet such that the crest of one, meets the trough of the other, their phases are opposite, resulting in what is known as destructive interference, since the two waves cancel each other at such points (Fig. 5.4b). Thus, an interference pattern may be regarded as a diagram of phase differences among waves. It is this interference pattern that forms the basis of holography that adds depth to otherwise flat images and make them spring to life.

The fact that light is a wave and not a stream of “corpuscles” was established in a classic experiment by Thomas Young in early 1800s. Young’s experiment, famously known as the double-slit experiment, consists of a light source whose light falls on a screen with two slits. The two slits now act as point sources of light from which light expands outwards in circular waves that interfere, constructively and destructively, forming intricate interference patterns. Light falling on another screen downstream shows a band-like pattern of intensity, with the bright and dark bands representing



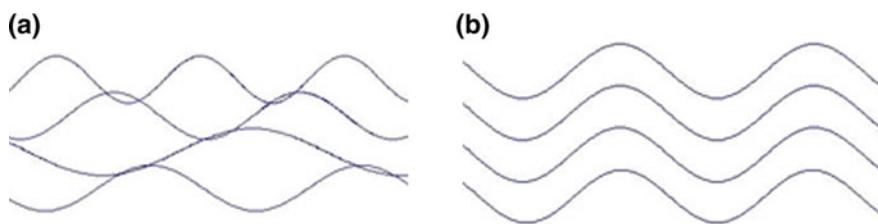
**Fig. 5.4** Constructive and destructive interference



**Fig. 5.5** A sketch of interference pattern obtained in Young's double-slit experiment

constructive and destructive interference, respectively. Thus, the interference pattern may be thought of as a diagram of the phase of the wavefront (Fig. 5.5).

Phase differences among waves have a lot of significance to our perception of three-dimensional objects and to the principle of the hologram. Phase difference has also a key role in our ability to localize a sound source, an auditory equivalent of perceiving depth and direction visually. Consider a sound that originated from your right. You can identify where it came from, even with your eyes closed, since the signal reaches your right ear first before it reaches the left one. Or, in the language of phases, the phase of the sound vibrations near your right ear leads that of the vibrations at your left ear. This is the principle used in stereo sound. By gradually varying the phase difference between the sounds that are played to the two ears, it is possible to control the perceived location of the sound source. Stretching the analogy to optics, when the light that bounces off a real object meets our eye, what it brings is not just a pattern of intensity (or amplitude) but also the phase. This phase is critical to our perception of a three-dimensional world of depth and shape. Note that phase is only one factor that defines our depth, or three-dimensional perception. Though there are others—use of two eyes to find depth, use of motion cues, or use of shading



**Fig. 5.6** **a** Incoherent waves, as in ordinary light. **b** Coherent waves, as in a laser

that arises from play of light and shadow, each providing its supply of depth-related information to the viewer—phase is primary.

The photofilm in a camera only records the intensity pattern, the 2D photo, of the light that enters the camera. A hologram is special in that it records the phase pattern. To this end, it uses a special optical device known as a laser. In ordinary light from an incandescent bulb, or a candle flame or the blazing sun, the phases are in a disarray, like the chaotic flow of vehicular traffic in Indian metros. Such light is said to be incoherent (Fig. 5.6a). In a laser, the waves move together, the phases are regular, like the *jawans* marching on the Republic day. A laser is a coherent light (Fig. 5.6b). In order to recreate the phase pattern of a light wave emerging from an object, we must first record the phase pattern on a film, the way the traditional photo is recorded on a film. This recording of phase pattern is known as a hologram.

To produce a hologram, the beam from a laser is split into two—one of these falls on the object of interest, while the other acts as a reference beam. When the beam that falls on the object, bounces off the object, its phases are altered in a manner that reflects the surface features of the object. This beam with altered phase pattern is now made to meet the reference beam and interfere with it. The resulting interference pattern, strongly reminiscent of the light and dark stripes in Young's experiment that we visited above, is recorded on a photographic film. The complex stripe pattern recorded on the film will have no resemblance whatsoever to the object that's being "holographed." It contains a code, a kind of a "bar code," that represents the object. Once the hologram is made, the original object may be visualized by simply shining a laser through the film, and seeing the image from the other side. When your vantage point changes as you view the image, you feel that you are seeing different sides of the object as though you are seeing the original object in three-dimensions.

Another interesting property of a hologram is the grace with which it responds to injury. A local damage to the hologram does not produce a local blank space in the scene; the entire image is degraded slightly. This happens because it is as though every point on a hologram contains information about the entire object. It is not a point-to-point map of the object, as is the case with the regular photograph. When Princess Leia chose to send her distress call across vast expanses of the universe in the form of a hologram loaded in R2D2's memory, she perhaps knew something of this magnificent robustness of holograms!

Karl Pribram is an Austrian-American who worked as a professor of psychology and psychiatry at Stanford University. Pribram was intrigued by holograms and saw a close resemblance between holograms and memories. Both have something “holistic” about them—in both the whole survives in every part. Pribram wondered if the mathematics of holograms can be used to unravel the substratum of memories in the brain. It is not that Pribram was the first to think of the hologram-memory connection. There were others who thought of it. But through the 60s and the 70s, through extensive writing and lecturing, Pribram rallied support and popularized the idea of a holographic theory of the brain.

Pribram focused his search on a specific kind of memories—visual memories. He began with a close study of the visual system to see if it has the machinery to construct something like a hologram of the images that enter the eye. Light entering the visual system begins its journey when it hits the cornea, a bulging transparent portion in the frontal part of the eye. This convexity of cornea enables it to behave as a lens and focus the incoming light beam to some extent. Further focusing occurs when light hits an actual lens, present a little behind the cornea, behind a little pool of clear liquid known as aqueous humor. The thickness of the eye lens is adjusted by special muscles that stretch the lens. This ability to vary the thickness of the lens gives the eye the ability to focus on near or far off objects. A thicker lens is used to focus on nearby objects, like a page of text; a thinner one is suitable for distant objects like a setting sun at the horizon.

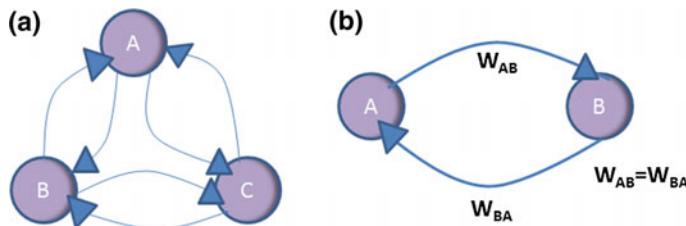
Light entering the eye then makes it beyond the lens and another body of colorless liquid beyond the lens, before it hits the retina, a place of intense action at the back of the eye. The retina has a layer of photoreceptors which convert the light signals into electrical signals which were transmitted onward to the brain. There are two kinds of receptors—rods and cones—which specialize in night vision and color vision, respectively. But the rods and cones do not directly project to the higher brain regions that process visual information. There are two further layers of neurons within retina, through which the electrical signals generated in the rods and cones through crisscrossing wires. The last layer of neurons in the retina, a layer of neurons known as ganglion cells, project to a central visual processing region in the brain called lateral geniculate nucleus, which is itself a part of a major sensory hub at the center of the brain—a structure known as thalamus. Inputs coming to the lateral geniculate nucleus from the ganglion layer further crisscross through the many layers in the thalamus before they are projected onwards to the primary visual cortex, crisply called V1, located at the back of the brain, in the occipital lobe. More crisscrossing occurs in this visual processing area, with neurons projecting to other neurons near and far, through complex networks of synapses. With so many intervening layers between the retina and primary visual cortex and intense crisscrossing of connections, it is quite likely that a spot in the image projects to several, or all, parts of the primary visual cortex. A stimulus presented to the eye might elicit electrical waves that originate in the retina and propagate all the way up the primary visual cortex. These waves might interfere with each other as they propagate over the cortical surface, forming intricate interfering patterns like those captured by a hologram. Perhaps these interference patterns are stored by the brain somehow and retrieved as memories.

These speculations about how memories might be coded in the brain did not go very far. Perhaps the holographic model is not the most natural metaphor that can explain the engram. One aspect of the hologram is certainly shared by the engram—the existence of the whole in every part. Local damage to the substrate (the brain or the holographic film) does not produce local destruction of memory. Memories are distributed. But the analogy perhaps ends there. The detailed mathematics of holography, and the process by which holograms are constructed—by setting up interference between a reference beam and a reflected beam—cannot perhaps be rigidly superimposed on the situation in the brain. Perhaps there are simpler ways of describing the distributed nature of memory representations. One such simpler method was discovered by John Hopfield.

## Recurrent Networks

John Hopfield, a theoretical physicist, who was at California Institute of Technology in the early 80s, was thinking of the problem of understanding memories in the brain. He went on to develop a model known as an Associative Memory, which is constructed by a network of our familiar McCulloch-Pitts neurons. However, compared to the perceptron or the MLP, which were also networks of McCulloch-Pitts neurons, the new memory network, popularly called the Hopfield network, has a distinct connectivity pattern. It may be recalled that neurons in a perceptron or an MLP are connected as a series of layers, with neurons in one layer projecting to the following layer. Therefore, information in such network always propagates in a specific direction, the “forward” direction, from the input layer to the output layer. Hence these networks belong to a larger class of networks known as feedforward networks. To reiterate the obvious, feedforward networks do not have “feedback loops.” There are no paths that begin in one neuron and return to the same neuron after a long tour through the network. Such networks were studied extensively not because the connectivity patterns in the brain have a precise, “feedforward” structure. Feedforward networks are simpler to analyze mathematically. There is a lot that can be explained about the brain even with these simpler networks. But brain has loops indeed.

Looping, or recurrent connectivity, is an extremely common motif in brain’s connectivity. There are small, micro loops within an area of a few hundred square microns, connecting groups of neurons doing similar function in a functional unit of the cortex called a cortical column. Or, a cortical area A can project to another cortical area B, while receiving reciprocal connections from B, forming a large  $A \rightarrow B \rightarrow A$  loop. Sometimes a cortical area can project to a subcortical area like basal ganglia, or hippocampus, and receive connections after several intermediate stages. Some loops can be even more extensive, spanning a significant stretch from, say, the highest regions of motor cortex, all the way down via the spinal cord to the limbs, and back, all the way up to the same motor cortical area, via several stopovers. Therefore, loops are definitely brain-like, but at the same time as much harder to analyze mathematically.



**Fig. 5.7** Structure of the Hopfield network. **a** Every neuron is connected to every other neuron

Hopfield noticed a network of McCulloch-Pitts neurons, with a recurrent structure—every neuron is connected to every other neuron—can be an elegant model of human memory, the associative memory (Fig. 5.7). The characteristic of an associative memory is best explained by contrasting it with the more familiar computer memory, more formally called an indexed memory. Memories in a computer are stored, like in a telephone directory, as one long list of numbered slots in which information is stored. Each slot has an address or an “index”, which is required to access the information content of the slot. Both the address and the content are coded in terms of bits—0s and 1s. Therefore, if an address has  $n$  bits, there can be totally  $2^n$  addresses. Thus, the number of addresses grows rapidly with the number of bits contained in the address. The strength of an indexed memory is its strict linear order, which also gives it an enormous size. Its linearity is also its weakness since to retrieve a piece of information stored in the memory, one has to know the address where it is stored. Otherwise, a brute force search will have to go over all  $2^n$  addresses.

The weaknesses of an indexed memory, and the need for an alternative, like the associative memory, can be easily brought out with the example of a telephone directory. The normal way of search in a telephone directory is with the help of the person’s name, typically the last name. If we know the last name, we can find the person’s first name, address and also telephone number. But sometimes we might require to identify the owner of a certain telephone number. Or we might desire to know the telephone number of a person who lives at a certain address. There is no obvious way of doing so in a voluminous, traditional telephone directory.

The difficulty with an indexed memory lies in the fact that there are a separate address and content. In a telephone directory, the last name is address, and the rest (first name, physical address, and telephone number) constitute the content. Alternatively, we would like the telephone number to serve as an “address” and be able to retrieve the remaining information as “content.” Thus we begin to notice that the distinction between the “address” and “content” in an indexed memory is an artificial one.

An alternative memory in which this sharp distinction between “address” and “content” disappears, is an associative memory. In an associative memory, there is only a “record” (e.g., a person’s information) with multiple fields (first name, last name, etc.). On a given occasion, one or more fields can together serve as an “address” while the remaining fields in the record form the content. But that is nearly how our

memories work. When you try to recall, for example, a special someone you have met long ago, it immediately brings to your mind the surroundings of that meeting, the time of that meeting, the events that may have led to that meeting, or any other relevant associations. Or if you try to recall the place of that meeting, other parts of the picture too are immediately recalled. Thus we do not seem to stock our memories against a long, mnemonic grocery list. We seem to remember one thing with the help of another. We remember things as a network of items, and not as twin-pleated lists of addresses and contents.

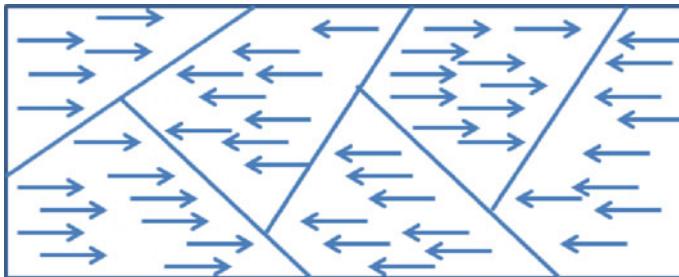
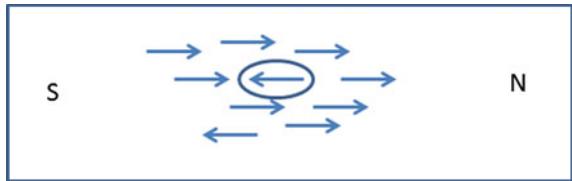
The recognition that an associative memory consists of a network of items, and therefore possibly be implemented by a network of neurons, led Hopfield to develop his network model of memory. But in writing down a mathematical formulation of such a network, what inspired him is not neurobiology, but a topic from statistical physics. Hopfield's fine insight consists of perceiving an analogy between the physics of magnetic systems and the brain.

The magnetism of a magnet has its roots in the motion of electrons in an atom. Moving charges constitute an electric current, which in turn produces a magnetic field that circulates around the direction of flow of the current. Particularly, if the charges flow in a closed loop, the magnetic field cuts the plane of the loop at right angles. Now it so happens that an electron in an atom undergoes two kinds of circular motions: one around itself, a property known as spin, and the other around the nucleus. These two forms of motion produce a tiny magnetic field. In some materials like the magnetic materials, actually ferromagnetic materials, if you want to be a stickler for terminology, all the tiny magnetic fields produced by all the electrons add up (and do not cancel out as it happens in non-magnetic materials) and produce a net magnetic field. Therefore, each atom in a magnetic material behaves like a tiny magnet. When a large number of these "atomic magnets" in a piece of magnetic material line up, the material exhibits a sufficiently strong magnetic field in a certain direction.

The process by which the tiny atomic magnets line up deserves a close examination since that is what had led Hopfield see the following analogy: atomic magnets → neurons, and magnetic material → brain. Each atomic magnet finds itself in a neighborhood of other atomic magnets with their own magnetic field pointed in a unique direction. If an atomic magnet finds itself in a neighborhood in which most other atomic magnets point in given direction (let's call it "north" for convenience), then their net northwards magnetic field acting on A tends to turn it also northwards, if it is not already turned in that direction (Fig. 5.8). Thus a local consensus builds up and expands until a considerably large number of atomic magnets in a small region of the magnetic material point in the same direction.

Thus a piece of ferromagnetic material contains small domains of magnetization, with the atomic magnets in the same domain pointing in the same direction. Adjacent domains might be pointing in quite different directions (Fig. 5.9). Atomic magnets in a given domain keep pointing in the same direction until acted upon by an external magnetic field. Thus they can be thought of holding on to "information." For example, a domain in which the field points toward "north" may signify a 1, or another pointing toward "south" may represent a 0. Thus a magnetic material may serve effectively as a memory, which is exactly what happens on a computer's hard disk. The head

**Fig. 5.8** An atomic magnet (circled arrow) is a neighborhood where most other atomic magnets are pointed north, forcing it to turn to the north



**Fig. 5.9** A depiction of “domains” in a magnetic material

that writes information on a hard disk actually produces a tiny local field and forces all the atomic magnets in a small domain to point in the same direction.

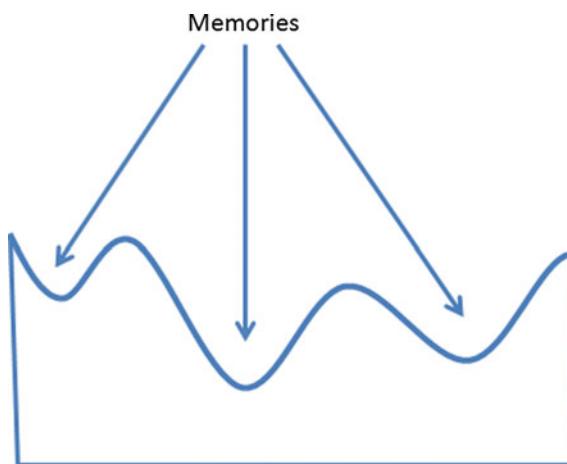
Let us try to abstract out a more general lesson from the above magnet and memory story. Let us just call the atomic magnets something abstract like “units” each of which is associated with a state (the direction of its field). The units interact among themselves and affect the states of each other. These interactions lead to certain collective states that are stable and do not change under small perturbations. Such stable states may be thought of as memories. The magnet/brain analogy comes alive if we think of the “units” as neurons, in either excited or resting states, interacting via synapses, influencing the states of other neurons.

Hopfield’s associative memory model is a network of McCulloch-Pitts neurons, in which each neuron is connected to every other neuron over a set of weights. Each neuron receives information about the states of all other neurons, weighted by the connecting weight values, and sums all such contributions. If the net input is positive, it updates its state to excited state (1), or it remains in resting state (0). This modified state is then sensed by other neurons. Thus neurons take turns flipping their states on and off. A key question that then arises is: does this flipping ever stop? Hopfield’s analysis tells us that the flipping does ultimately stop when a certain assumption about the weights is made.

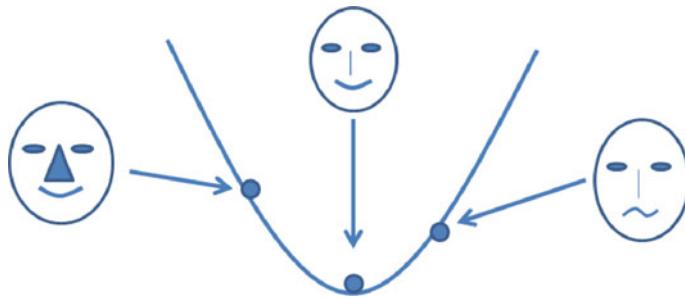
In order to make sure the flipping behavior stops ultimately, and the neurons fall into a stable state, Hopfield makes a slightly restrictive assumption about the weights: they have to be symmetric. That is, the weight on the synapse from neuron A to neuron B, is the same as the weight on the return connection from neuron B to neuron A. Although this assumption is sometimes justified by the fact that reciprocal connections are a common phenomenon in the cortex, it is introduced more for

mathematical conveniences it affords. When the weights are symmetric, Hopfield proved that there exists a nice, convenient function—which he calls the “energy” function, definitely a hangover from the magnetic system metaphor, which assures that the network dynamics is stable. Each collective state (the set of all individual states of the neurons) of the network is associated with an energy value. As the neurons flip their states, the corresponding energy goes to a minimum. That is the network searches and settles in states of minimum energy, like a ball tossed over a rolling terrain, rolls down the nearest trough and settles there. Another important feature of Hopfield’s energy function is that it does not have a single unique minimum energy state. The energy function of a Hopfield network can have a large number of minima, called local minima. The evolving network state can settle down in any of those local minima. Hopfield visualized these local minima of the energy function as memories of the network (Fig. 5.10).

Let us consider how exactly these local minima work as memories. Recall that every state in the network is a pattern of 1’s and 0’s, which are the states of the neurons in the network. This pattern of bits could be thought to represent a “memory” like, say, that of an image, since in this digital age we are aware that a great variety of things—songs, images, videos—are simply streams of bits. For example, a local minimum of energy in Fig. 5.11, could represent an oversimplified “smilie”. Nearby points on the energy function, then, could be imagined to be small variations of the smilie—with a twisted smile, a winking eye, or a stub nose. Imagine the network in a state corresponding to an “eye wink.” This state is not stable since the neighboring states have lesser energy. The network rolls down from its “eye wink” state to the normal smilie with two, bright, unwinking eyes at the bottom of the well. A similar return to the normal smilie will be witnessed when the state is perturbed toward a “twisted smile” version, or the one with a “stub nose.”



**Fig. 5.10** Memories as the minima of the energy function of the Hopfield network



**Fig. 5.11** At the bottom of the attractor there is a stored memory of a “smilie.” States in the neighborhood of the perfect stored memory are distorted versions of the memory—like a smilie with a stub nose or one with a twisted smile

Let us expand the picture a little and consider an energy function with two local minima, corresponding to two patterns A and B. If the network begins in the neighborhood of pattern A, it rolls down and settles in pattern A. Similarly from the neighborhood of pattern B, it rolls down to B. This rolling down to a pattern is thought of as a process of “memory retrieval.” The state at the bottom of the well is the stored memory, the one that is retrieved. Neighboring states are the “cues”—like the eyes in a photograph that cue recall of an entire face—that help retrieve a corresponding memory. Each memory is cued by only certain cues, the relevant ones.

We described the process by which the Hopfield network evolves toward a stored state from the initial state that is a cue to the stored state. We also mentioned that the stored state is a stable state, a local minima in the energy function. But how is the landscape of the hills and valleys—the local minima—of the energy function shaped that way? Naturally, a network with random weights does not retrieve a picture of Tajmahal! How do we store the patterns we would like to store in the network? How do we make sure that the network’s stable states, or local minima, are exactly the memories we would like to store and retrieve?

Hopfield’s solution to this problem is a rule that uses the stored patterns to set the network’s connections. The rule is best described by taking a sample pattern which needs to be stored in a network. Let us try to store,  $S = \{1\ 0\ 1\ 1\ 1\}$ , a pattern with 5 bits, which requires a network with five neurons. Now a five-neuron network has  $5 \times 5 = 25$  weights, with each neuron connected to every neuron, including itself. The rule for storing the pattern may be stated as follows:

The weight connecting  $i$ th neuron and  $j$ th neuron = ( $i$ th bit of  $S$ )  $\times$  ( $j$ th bit of  $S$ )

Ex1:  $w_{23}$ , the weight connecting neurons #2 and #3, equals  $S_2 \times S_3 = 0 \times 1 = 0$ .

Ex2:  $w_{15}$ , the weight connecting neurons #1 and #5, equals  $S_1 \times S_5 = 1 \times 1 = 1$ .  
Similarly, all the 25 weights are calculated.

In a more general form, the rule for storing a pattern,  $S$ , with  $n$  bits, in a  $n$ -neuron Hopfield network, is expressed as,

$$W_{ij} = S_i \times S_j$$

The above rule may be summarized in plain English as follows. If the states of  $i$ th and  $j$ th neurons are simultaneously “high” ( $S_i = 1 = S_j$ ), the connection between them is also “high” (1). Even if one of the neurons ( $i$ th or  $j$ th neuron) is 0, the connection between them is “low” (0).

This rule has an interesting history. First of all, the rule can be derived using simple high-school level mathematical reasoning. All that is required is, given a pattern  $S$ , to find the weights that minimize the energy function associated with the network. We simply need to construct a quadratic function of  $S$  with an appropriate form. But the rule is also reminiscent of a cellular level mechanism of memory proposed by Donald Hebb, around the middle of the twentieth century. Therefore, the rule is often called the Hebb’s rule.

Donald Hebb (1904–1985) was a Canadian psychologist who had an enormous influence on behavioral and computational neuroscience. Inspired by the writings of stalwarts like James, Freud, and Watson in his formative years, he took to a career in psychology. As a post-doctoral fellow, he worked with Wilder Penfield at the Montreal Neurological Institute where he studied the effect of surgery and lesions on memory and behavior. Like his predecessor Lashley, Hebb concluded that memory in the brain has a distributed representation. He was preoccupied with the question of how memories are coded as synaptic strengths. His experiences led him to create a “general theory of behavior that attempts to bridge the gap between neurophysiology and psychology.” He published the theory in 1949 in a book titled: “The Organization of Behavior: A neuropsychological theory.” The book was received with great enthusiasm when it appeared and was described, along with Darwin’s “Origin of the Species,” as one of the two most important books in biology.

Hebb’s fine insight lies in seeing that ongoing neural activity could control synaptic strength. As one of his oft-quoted statements goes

When an axon of cell A is near enough to excite B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased. (p. 62)

The above statement basically says that if a neuron A repeatedly succeeds in firing B, the connection from A to B must be strengthened. In popular parlance, this idea is sometimes expressed as: “neurons that fire together wire together.”

Starting from his powerful rule, Hebb went on to postulate that in a network in which synapses obey Hebb’s rule, neurons cluster into cell assemblies. Neurons that tend to respond to a common stimulus, tend to fire together and therefore wire together. Due to this strengthening of within-group connections, neurons that respond to a common stimulus also tend to excite each other, thereby forming a cell assembly.

It is quite interesting that Hebb’s speculations, having their roots in psychology and behavioral science, strongly resonate with the storage formula of Hopfield network, which is derivable from simple mathematical considerations. Perhaps this is yet another example of what physicist Eugene Wigner called the “unreasonable efficacy of mathematics.” But Hopfield’s storage rule can only store a single pattern. Each of

us stores a large number of memories of many varieties—many faces, many names, many routes, and many places. Therefore, realistically, the Hopfield network must be able to store multiple patterns. A slight generalization of Hebb's rule we encountered above allows us to store multiple patterns. If we wish to store two patterns,  $S^1$  and  $S^2$ , to begin with, the rule is

$$W_{ij} = S_i^1 \times S_j^1 + S_i^2 \times S_j^2$$

The weight connecting  $i$ th and  $j$ th neurons,  $w_{ij}$ , now has two components, one from each stored pattern. More generally, to store  $P$  patterns,

$$W_{ij} = S_i^1 \times S_j^1 + S_i^2 \times S_j^2 + \cdots + S_i^P \times S_j^P$$

Using the last-mentioned rule multiple patterns can be stored and retrieved securely from a Hopfield network. Most importantly, information that codes for the patterns is distributed all over the network. It is present together in each one of the network connections. There are no separate connections that code for individual patterns. The parallel and distributed processing seen in an MLP is again encountered in the memory dynamics of Hopfield network.

A natural question that emerges is: how many patterns can be stored in a Hopfield net using the above formula. Note that the formula itself is general, and does not put any restriction on the number of patterns that can be stored. But Hopfield discovered in his early studies with this network, that as more and more patterns are stored, the network gets, in a sense, overloaded, causing increasing errors in retrieval. These errors grow rapidly when a critical number of patterns are stored. Using numerical simulations, Hopfield noticed that in a network of " $n$ " neurons, about  $0.14n$  patterns can be stored. Thus about 14 patterns can be stored in a 100 neuron network. Subsequently, this limitation has also been explained mathematically. This limitation on the number of patterns that can be stored in a Hopfield network is known as its memory capacity.

A 14% capacity seems to be seriously limiting, considering that in indexed memory of a computer an  $n$ -bit address can be used to code for  $2^n$  items. But from the standpoint of brain's infrastructure, with its 100 billion neurons, 14% is not all that restrictive—it stands for a storage space of about 14 billion patterns, a superhumanly massive storage space by any stretch of imagination. But it is not right to apply the Hopfield's memory capacity result directly to the brain since brain is not a fully connected network. Each of the 100 billion neurons is connected to only about 1000 or 10,000 neurons on an average. There are mathematical extensions of Hopfield network in which each neuron is connected to a small fraction of neurons in the network. Since such a network will have fewer connections, than a fully connected Hopfield network, it appears at the first glance that such partially connected networks will have lesser storage capacity. However, it turns out that, with appropriate types of connectivity, even such partially connected networks could show counterintuitively high memory capacity. But we refrain from proceeding along this line of thinking for the moment for several reasons. First of all, it is a folly to think of the brain as

one large Hopfield network whose sole function is to store patterns. A specific set of brain structures form the brain's memorizing architecture. Sensory information that enters into the brain, flows through specific pathways, crosses a series of stages before it forms part of our memories, specifically our long-term memories. It would be hasty to apply the lessons learnt from Hopfield's model to the cerebral context without considering the precise substrates of memories in the brain.

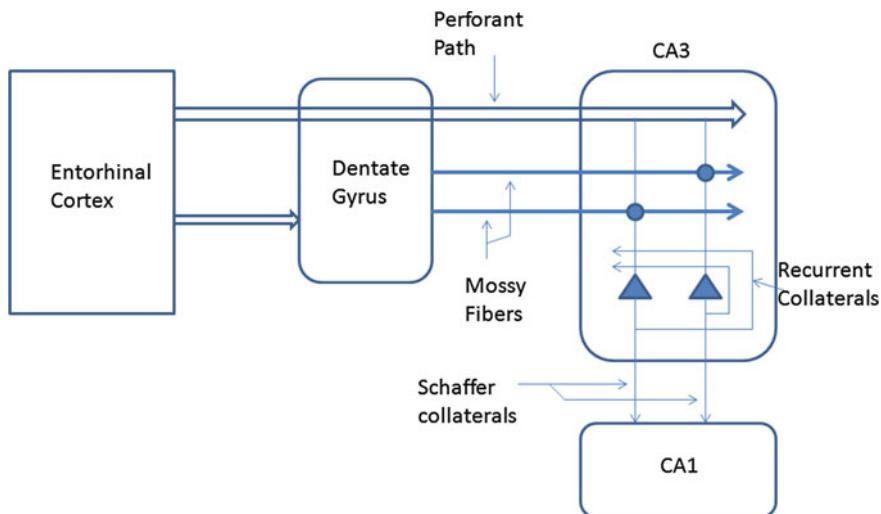
Hopfield network is a remarkable starting point that leads us on a very fruitful line of study about brain's memory mechanisms. The Hopfield network with its richly connected McCulloch-Pitts neurons, its energy function, its local minima that stand for stored memories, offers a convenient framework to think about memories in the brain. Next time you recall the face of an old friend when you hear her name mentioned, you could almost hear the neurons in your brain flashing spikes, vigorously flipping states between excitation and rest, pushing toward a massive brain-wide neuronal consensus, and recreating a brain state that produces the experience of actually gazing at your friend's face, or a picture of the same in fond remembrance. However, unfortunately, brain scientists would be having a great time if life were to be that simple. A lot of brain's recalcitrant secrets could have been solved in a few years if simple mathematical models could be quickly applied to real brain situations to make sense of the stubborn perplexities of the mental and neural world. Therefore, at this point, it would be a worthwhile task to consider some relevant data from neurobiology of memory.

Let us begin with a simple question: is Hebbian mechanism for real? Hebb's rule plays a key role in the formulation of Hopfield network. It certainly has the strength of mathematical justification and the support of Hebb's brilliant speculations. But are real synapses Hebbian? Let us consider some crucial experiments in this direction.

## Synapses that Memorize

A lot of knowledge of synapses that exhibit long-lasting changes has come from studies of synapses located in a part of the brain known as hippocampus. The word comes from the Latin term which means "seahorse" thanks to the obvious resemblance of this brain structure to the aquatic creature. Located inside the temporal lobe of the brain, hippocampus also involved in memory formation. Damage to hippocampus is also known to cause serious impairment in the ability to acquire new memories. (More on this later). Synapses in hippocampus are found to show long-lasting changes under special conditions of stimulation.

In these stimulation experiments, typically, a slice of the brain is taken in such a way that the slice cuts through and exposes a section of hippocampus. The schematic of Fig. 5.12 shows three pools of neurons in hippocampus: dentate gyrus, CA3, and CA1. Hippocampus receives its inputs from entorhinal cortex, a part of the cortex located on the medial (inner) side of the temporal lobe. Fibers from entorhinal cortex project to the first stopover in hippocampus, the dentate gyrus. Some fibers, which constitute the perforant pathway, also cut through directly to the CA3 region. Neurons



**Fig. 5.12** Schematic of key connections within hippocampus

of dentate gyrus also project to CA3 through fibers called mossy fibers. CA3 neurons further synapse onto neurons of CA1 via fibers known as Schaffer collaterals.

Several synapses in the hippocampal network briefly described above are capable of showing lasting changes. We are now familiar with the idea that the “strength” of a synapse consists of the post-synaptic potential it produces in response to an action potential that arrives at the synapse from the pre-synaptic side. Modification of the synaptic strength, a property known as plasticity, is thought to underlie most learning and memory-related phenomena in the brain. A synapse may be said to have “changed” when it shows a significant change in its “strength” in response to a pre-synaptic action potential.

In an experiment performed by Robert Malenka and coworkers, intracellular recordings are taken from neurons of CA1 region, when the Schaffer collaterals are electrically stimulated. Pre-synaptic stimulation is usually done with a high-frequency volley of pulses for best results. Stimulation electrode is also placed in CA1 neurons to inject a positive current that depolarizes the membrane voltage. This joint stimulation of both pre- and post-synaptic sides, even though only for about 15 min, is shown to have a dramatic effect on the synaptic strength. After the stimulation session, when the pre-synaptic terminal is stimulated by a low-frequency stimulation, much larger PSP's are recorded compared to the PSP levels found in the synapse before the stimulation. Moreover, these changes last for several hours. When studied in a live animal (*in vivo*) the changes were found to persist for several weeks. Another interesting aspect of this change is that when pre- and post-synaptic stimulation was done, not simultaneously, but alternatively, there was no appreciable change in the synaptic strength. Similarly, such changes were not observed when only the pre-synaptic terminal was fired, but the post-synaptic side was not depo-

larized, or when it was actually prevented from depolarizing by injecting a negative current. Thus, simultaneous activation of the pre- and post-sides seems to be a crucial requirement for the production of lasting synaptic changes. This type of synaptic change is known as Long-term Potentiation (LTP). LTP was first observed by Terje Lømo in 1966 in his studies on the synapses of perforant pathway to dentate gyrus. Subsequently, LTP was found not only in other synapses of hippocampus, but in many other brain areas. Thus LTP is a convincing vindication of Hebb's speculation that synapses that are activated by a pre-and-post stimulation tend to be strengthened.

Having come this far down the "memory" lane, and having encountered a possible cellular mechanism of memory, let us take a walk a few further steps down, and look at some of the key molecular players in the induction of LTP. We begin by mentioning that the LTP experiments described above were performed in hippocampal synapses where the neurotransmitter is glutamate, an excitatory neurotransmitter. Glutamate has two important ionotropic receptors known as—AMPA ( $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole-propionate) and NMDA (*N*-Methyl-D-Aspartate). Glutamate released from the presynaptic terminal activates AMPA, which opens an associated  $\text{Na}^+$  channel, causing an influx of  $\text{Na}^+$  ions into the post-synaptic side, leading to generation of an Excitatory Post-Synaptic Potential (EPSP) or a positive deviation in the post-synaptic membrane potential. A couple of ways in which a synapse can be strengthened is by increased post-synaptic receptor density or by increased release (or release probability) of pre-synaptic neurotransmitter when hit by an action potential. But what triggers this change? Note that LTP is induced by a double event: glutamate release and sufficiently large post-synaptic depolarization. What is the molecular machinery that can detect this double event? Interestingly the other key glutamate receptor, NMDA, seems to work the charm necessary for LTP induction.

Unlike AMPA which is essentially a ligand-gated receptor, NMDA is both ligand-gated and voltage-gated. Since it is ligand-gated, it is capable of sensing the glutamate signals from the pre-synaptic side, and since it is voltage-gated, it is able to detect depolarization event on the post-synaptic side. Thus, NMDA activation is a crucial event that results in a cascade of molecular events on both pre- and post-synaptic ends, resulting in induction of LTP.

Subsequently, LTP has been discovered in other brain regions like the cortex, cerebellum (a prominent structure in our motor system), and amygdala (a structure that processes our fear response). In fact, Robert Malenka, an LTP researcher, believes that LTP may occur at all excitatory systems in mammalian nervous systems. There are also other variations of LTP. For example, in hippocampus itself, in the mossy fiber pathway, there exists a non-Hebbian type of LTP which can be induced by high-frequency pre-synaptic stimulation alone without the concomitant post-synaptic depolarization. There is yet another kind of LTP, known as the anti-Hebbian LTP, which can be induced by depolarizing the pre-synaptic terminal while hyperpolarizing the post-synaptic terminal. Thus, in this case, the synapse is strengthened when pre- alone, but definitely not the post-synaptic side, gets activated.

What the preceding discussion tells us is that there are a variety of cellular and synaptic mechanisms that can retain traces of activation histories of connected neuronal pairs in some form of synaptic strength. A key requirement for the Hopfield

network to work, viz., the Hebbian mechanism, is now known to exist in the form of LTP. It is also interesting that synapses that show LTP are present in hippocampus. But our memory story is far from complete. We have only a necessary cellular mechanism. Where is the evidence that such cellular mechanisms operating in a larger network of real neurons (and not just a mathematical network, Hopfield or otherwise) are responsible to phenomena of learning and memory at behavioral level? As a first step to answer this question, let us begin with the story of hippocampus and its link to short-term memory loss.

## A Scratchpad of Memory

One of the first hints regarding hippocampus and its connection to memory came with the study of a patient known as Henry Gustav Molaison, famously known as HM in neuroscience literature. Henry was an American born in 1926 in the state of Connecticut. A bicycle accident at the age of seven was supposed to be the cause of recurrent epileptic attacks Henry suffered for many years. When a surgeon was consulted in 1953, it was discovered that the source of Henry's seizures was localized to his medial temporal lobes, a brain region that houses hippocampus (actually two hippocampi, one on either side of the brain). A surgery was performed the same year and parts of Henry's medial temporal lobe, on either side, were removed.

Although the surgery successfully reined in Henry's epilepsy, a new problem emerged. Henry is now unable to store memories for an extended period of time. He suffered from amnesia, not of one kind but two: anterograde amnesia and retrograde amnesia. He had strong anterograde amnesia, an inability to store new memories. He also had a milder form of retrograde amnesia, which refers to an inability to recall old memories. Particularly, he had poor memory of events that immediately preceded his surgery, but remembered older past. He had difficulty in remembering events that happened a few years before his surgery but remembered his childhood events. But since he was unable to create new long-lasting memories, he literally lived in the present. He was able to retain information, however, for short periods of time. His working memory was intact. His ability to recall numbers presented a little while ago was indistinguishable from normal subjects. Another kind of memory he was able to acquire is motor skill. He was able to learn, for example, how to copy a figure by looking at its mirror image.

Thus, Henry's impairment seems to be limited to inability to form new long-term memories, though memories if they can be so called, of skilled behaviors are intact. Since the surgery destroyed parts of his hippocampi bilaterally, it seemed reasonable to assume that hippocampus is responsible for the ability to store new memories.

Let us pause for a moment to take our current conceptual bearings. We began by presenting the Hopfield network as a reasonable framework to describe the kind of neural dynamics in the brain that might support memory storage and retrieval. It might perhaps be too naïve to expect the precise mathematical formulation of this model to explain the real memory dynamics of the brain. The Hopfield model may

be regarded as a conceptual basket that can be suitably applied to study memory in the brain. Since the model combines the McCulloch-Pitts neuron, a plausible neuron model, and Hebbian learning, our confidence in the model is further enhanced. We proceeded with this line of study to find direct neurobiological evidence to various features of Hopfield's model. We found that Hebbian learning is not a mere hypothesis but actually occurs in real synapses, and more importantly, it occurs in a special brain structure—the hippocampus—a place where memories seem to be parked for a short while before they are moved to longer term stores.

But the picture we have so far is quite a coarse one. If we view the problem at a higher resolution, we will have many questions to answer. What is it about the hippocampus that enables it to play the role of a memory store? Presence of Hebbian synapses is not enough and is perhaps also irrelevant if it does not have an appropriate network structure. Hebbian synapses in a feedforward network, for example, do not give us a Hopfield network and attractor dynamics. Hebbian synapses co-exist with a richly recurrent connections in a Hopfield network, a combination essential for its memory-related dynamics. Does hippocampus have such rich recurrent connectivity pattern?

It is interesting that neurons in CA3 part of hippocampus indeed have rich internal connections. In any brain region, neurons do make connections with nearby neighbors. But such recurrence is particularly enhanced in CA3. A single pyramidal neuron in CA3 region in rat, an animal extensively studied in hippocampus research, receives about 4000 synapses from entorhinal cortex via the perforant pathway but receives about 12,000 connections from other CA3 neurons. Effectively each CA3 neuron receives inputs from 4% of total CA3 pyramidal neurons. This level of recurrent connectivity may seem to be small compared to 100% recurrence found in a Hopfield network, but it is orders of magnitude higher in degree than that found elsewhere in the brain.

Therefore, by finding the recurrence we were looking for among the neurons of one of the key hippocampal regions, we added one more feather to the cap of neurobiological evidence for hippocampus to support Hopfield-type dynamics. We do have several pieces of evidence now. But the pieces are far apart and do not yet form a seamless conceptual fabric. For example, how are we sure that hippocampus, with its richly recurrent anatomical microstructure, its Hebbian mechanism, is actually responsible for memory and learning at behavioral level? The only coarse link we have with hippocampal dynamics and memory-related performance at behavioral level is the example of hippocampus-damaged patients like HM. More direct evidence is desirable.

For a more direct evidence linking hippocampal dynamics with behavior, researchers resorted to study animal brains, which offer more freedom for scientific exploration, than human subjects. Rats have been the select choice of hippocampus research for a long time, thanks to their inimitable spatial learning abilities, a strength that also seems to have its roots in hippocampus. To this end, Richard Morris and his colleagues studied the performance of rats in a water maze. In this experimental setup, typically, a rat is dropped at a random location in a large pool of water. The rat has to swim for its life and arrive at a platform located somewhere in the pool. In

some cases, the platform is visible, much to the relief of the drowning animal, but in others, the platform is kept slightly submerged. The water is made slightly muddy so that the platform is not visible through the water. The animal has to randomly explore the pool and search for the hidden platform. Once it arrives at the invisible platform, it is removed and repeatedly dropped at random locations. Even though the platform is hidden, a rat is usually able to find the platform faster with growing experience, since it is able to somehow encode the implicit location of the hidden platform, with respect to the spatial model of the world that it creates in the brain. The success with which it learns the task is measured by the time the animal takes to find the platform from a random initial location. This time taken to find the platform, known as the escape latency, is found to decrease as the animal acquires more and more exposure to the pool and its surroundings. The relevance of this maze learning story to our hippocampus story is that an intact hippocampus seems to be crucial even in this form of spatial learning.

To test the relevance of an intact hippocampus to the water maze learning, Morris and his colleagues chemically tampered with neurotransmission in hippocampus using a substance called Amino-5-phosphonopentanoate (**AP5**). AP5 is a NMDA receptor antagonist, which blocks formation of LTP. Its effect has been first verified in a brain slice experiment, where it was found that the strength of LTP decreases almost monotonically with increasing AP5 concentration. The parallel effect of AP5 on maze learning performance of the animal is quite revealing. Escape latency increased almost monotonically with increasing AP5 concentration. Thus an effect that blocked LTP at molecular level is also found to block memory and learning at behavioral level, as manifested in water maze learning abilities. This result shows that what is seen as LTP at molecular level is responsible for learning at behavior level.

Various pieces of neurobiological evidence, that support presence of Hopfield network-like machinery in hippocampus, subserving its role as a storehouse of short-term memory, are now falling in place. But there is still an important issue we have only glossed over previously. Let us reconsider storage and recall processes in the Hopfield network. Storage of patterns is done using the Hebbian formula, for storing single and multiple patterns alike. This is done as a separate one-shot step. Once the connections are thus computed, the network can be used to recall stored patterns. But such a state of affairs is rather artificial when seen from the perspective of neurobiology. In a mathematical model, we may have a separate storage stage, and a recall stage. But in the real brain, in the real hippocampus, what determines how an input that enters the portals of hippocampus is actually processed? Is it taken as a new pattern that has to be stored by appropriate modifications of the recurrent connections of CA1 or CA3? Or is it interpreted as a cue to a pre-existing memory, with the hippocampus responding with a completed pattern? What is the signal or the mechanism by which such a switch, assuming it does, takes place in the hippocampus?

We now have a situation in which a good computational model constantly anticipates and brings together the relevant facts from neurobiology, and helps weave a progressively coherent picture of the functioning of hippocampus. One clue was found in the anatomical connections among the various modules within hippocam-

pus. We may recall here, that the port of entry into the hippocampus is the entorhinal cortex, projections from which penetrate deeper into the hippocampus via two pathways. There is the perforant pathway, that projects directly on CA3 from entorhinal cortex, while a secondary pathway projection onto dentate gyrus first before projecting further onto CA3 via a pathway known as mossy fibers. Thus entorhinal cortex has access to CA3 via two distinct pathways—perforant pathway and mossy fiber pathway—both of which are very different in their action on their target. Mossy fibers are sparse but physically larger and stronger. The direct entorhinal inputs are more numerous but weaker. It is sufficient to stimulate a small number of mossy fibers to activate a CA3 pyramidal neuron, while it takes hundreds and thousands of perforant pathway fibers to achieve the same. It is quite tempting to visualize that the mossy fibers are used to imprint a pattern of activity arising in the entorhinal cortex onto CA3, which is probably necessary for pattern storage within CA3. On the other hand, a weaker stimulus, which denotes a cue, a hint, a prompt to recall a pre-existing pattern, may be carried by the weaker perforant pathway fibers.

Thus presence of an appropriate anatomical circuitry that can possibly offer the “dual route” necessary for storage and recall is in itself exciting but far from being the complete story. First, how or what does the switching between these two pathways, depending on the purpose for which hippocampus is used in a given instance—storage or recall? Second, the processes that occur within CA3 must be different between storage and recall. During storage, the recurrent synapses among the CA3 pyramidal neurons must be temporarily rendered more plastic, than during the recall stage, so that the strong input coming via the mossy fibers is lapped up and memorized by the CA3 network. Furthermore, there is yet another distinctive change that must take place to securely distinguish the operations underlying storage and recall stages. During recall, not only must the recurrent connections of CA3 remain fixed, non-plastic. They must also be more dominant than the inputs that come from the entorhinal cortex. Note that information flows in opposite directions during storage and recall: from the cortex into the hippocampus, via entorhinal cortex during storage, and from the hippocampus via entorhinal cortex back to cortical targets, during recall. Thus during storage, CA3 must be in “listening” mode, alert to the inputs from the entorhinal cortex, while shutting out its own inner whisperings generated by its internal recurrent connections. During recall, CA3 must be in “saying” mode, freely allowing expression of the activity of CA3 neurons, while shutting out the inputs coming from entorhinal cortex (Fig. 5.12). Thus a lot of precise and timely routing of information in and out of hippocampus has to take place for this structure to serve as a memory store. It is not enough to have a vague idea that LTP occurring in hippocampal synapses somehow supports our memory-related operations at behavioral level. There is a more difficult question that remains to be answered. What is the mechanism by which that precise routing of information takes place in the hippocampus?

## Acetylcholine and Hippocampal Machinery

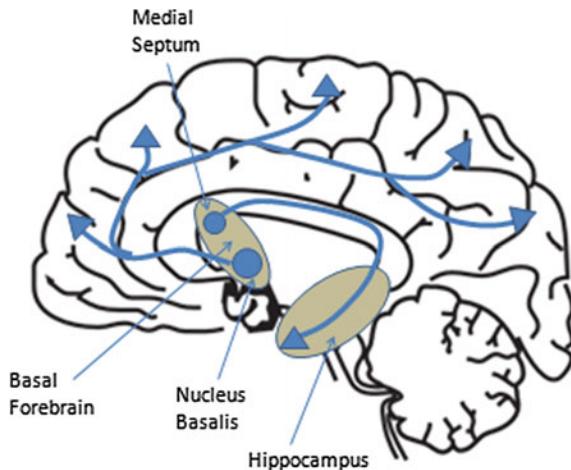
Acetylcholine is a neurotransmitter that carries commands from motor nerves to skeletal muscle resulting in muscular contraction. Since acetylcholine acts at the meeting point of the nervous system and the muscle—a region analogous to synapses among neurons, known as the neuromuscular junction—it plays an important role in expressing motor commands. Disruption of acetylcholine transmission can, therefore, be lethal, as it happens when botulinum toxin is injected, even in minuscule doses, into the body. Botulinum disrupts acetylcholine transmission by blocking the vesicles in axonal terminals from releasing acetylcholine. Acetylcholine is also the neurotransmitter used by the parasympathetic nervous system, a branch of our autonomic nervous system. The parasympathetic nervous system regulates the “rest and digest” activities in the body. Since the heart rate slows down under resting conditions, the action of parasympathetic system on the heart is to slow it down, an action that is mediated by acetylcholine. Therefore, it is obvious what damage can ensue by disruption of transmission of acetylcholine.

But what was described in the previous paragraph pertains to the functions of acetylcholine in the peripheral nervous system, the part of the nervous system that closely interfaces with the body. The other part of the nervous system, the non-peripheral or the central nervous system, consisting of the brain and spinal cord, is housed securely in a bony cage: the brain in the cranial vault and the spinal cord in the vertebral column. Within the central nervous system, acetylcholine dons its new avatar as a neuromodulator instead of a neurotransmitter. While a neurotransmitter carries a neural signal from the pre- to the post-synaptic terminal, a neuromodulator modulates the signal carried by the neurotransmitter by amplifying or attenuating it.

Another revealing aspect of neurons that release neuromodulators is an anatomical attribute of these neurons. Neurons that release neuromodulators are typically part of what is known as diffuse projection systems. These are usually organized as small neuronal clusters that send out rich and diffuse projections to large portions of the brain. There are four prominent diffuse projection systems: (1) the dopamine system, (2) the norepinephrine system, (3) the serotonin system and (4) the acetylcholine system. Since each of these systems can influence large parts of the brain, disruption in transmission of any of these systems can cause neurological and neuropsychiatric disorders. For example, disruption of dopamine transmission is implicated in Parkinson’s disease, a neurodegenerative disorder associated with debilitating motor impairments. Similarly, disruption of the acetylcholine system is associated with Alzheimer’s disease, an incapacitating disease of our memory function.

The link between acetylcholine and memory has its roots in the fact that one of the several targets to which acetylcholine-releasing, or cholinergic, neurons project includes hippocampus. Cholinergic neurons are located in several areas of the brain including the basal forebrain. Within the basal forebrain, one cluster of neurons, named the nucleus basalis projects to widely distributed targets all over the cortex. The other cluster, known as the medial septum, projects to the hippocampus, a projection that is sometimes known as the septohippocampal projection (Fig. 5.13).

**Fig. 5.13** Cholinergic pathways in the brain. The projection from medial septum to hippocampus is the septohippocampal pathway



Blocking the signals of the septohippocampal projection is known to disrupt memory function. This blocking can be done pharmacologically, using, for example, a cholinergic antagonist like scopolamine. Subjects who were given a list of words to memorize were administered with scopolamine. When tested for recall after some time, those who were given the drug fared much worse than control subjects. In extreme cases, scopolamine subjects did not even have memory of taking the test!

The opposite effect is seen in the case when an agonist of acetylcholine is injected into monkeys performing a simple memory task known as the Delayed Non-Match to Sample. In this task, the animal is originally shown a novel pattern. After a short delay, when the animal is shown a set of objects that includes the original object, it has to choose a novel object other than the original object. Monkeys that were injected physostigmine, an agonist of acetylcholine, in moderate doses, performed better than animals that did not receive the drug. Studies of this kind clearly establish the strong relevance of acetylcholine signaling in hippocampus to memory function.

But how do we understand the precise role of acetylcholine in hippocampus? Is it a general facilitator of hippocampal activity, so that reduction in acetylcholine levels generally attenuate hippocampus-related memory function? It turns out that the effect of reduced acetylcholine is simply identical to hippocampal lesion. This was observed on a task known as eyeblink conditioning. To put in very general terms, conditioning, in psychology refers to learning to associate stimuli to responses. More precisely, if a stimulus  $A_1$  produces a response  $B$  naturally in an animal, teaching the animal to respond with the same response  $B$ , to a new stimulus  $A_2$ , is known as conditioning. Thus the new stimulus-response ( $A_2 \rightarrow B$ ) acquired, in some sense, rides over the pre-existing stimulus-response ( $A_1 \rightarrow B$ ).

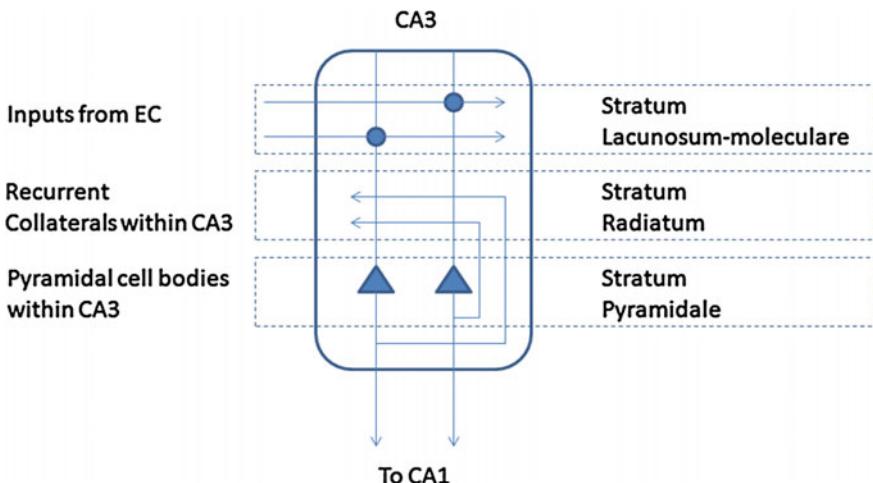
One form of conditioning, known as classical conditioning, is also known as Pavlovian conditioning, after the Russian psychologist who performed pioneering experiments with dogs on this form of conditioning. When a plate filled with food, like meat, is presented to a hungry dog, its natural response would be to salivate.

In Pavlovian conditioning, the presentation of meat is preceded by another stimulus like, say, the sound of a bell. On repeated pairings of the two stimuli—the sound of the bell followed by presentation of food—the animal learns to predict arrival of the food the moment it hears the sound of the bell and begins to salivate, even before the food is presented. The first stimulus, food, is known as the Unconditioned Stimulus (US), and the natural response of the dog to the same is known as Unconditioned Response (UR). The new stimulus which the dog learns to associate with the response (salivation), is known as the Conditioned Stimulus (CS) because it is a result of conditioning, or special learning, and was not present earlier.

Like the dog and salivation, another instance of classical conditioning is the eye-blink conditioning. In this experiment, a mild puff of air (the unconditioned stimulus, US) is blown on the cornea of the subject, whose natural response (unconditioned response, UR) is to blink. When the US is repeatedly preceded by a conditioning stimulus (CS) like a brief, audible tone, the subject learns to respond to the CS by blinking. Interestingly, though damage to hippocampus did not affect the development of eye-blink conditioning, disruption in acetylcholine signaling did. For example, eyeblink conditioning developed more slowly when scopolamine (acetylcholine antagonist) was administered. Similarly damage to medial septum, the origin of acetylcholine projections to hippocampus, also produced impairment in development of eyeblink conditioning. This brings the question of the precise role of acetylcholine in hippocampus into sharp focus.

At the end of the previous section, in our effort to superimpose the associative network theory onto the machinery of hippocampus, we felt the need for a mechanism that would switch between two modes of functioning: storage, in which the hippocampus is in the “listening” mode, and the internal connections of CA3 are plastic, willing to absorb the inflowing information, and recall, in which the hippocampus is in “saying” mode, in which the internal connections of CA3 are frozen, and the state of the CA3 network is read out.

Michael Hasselmo and coworkers have suggested that acetylcholine plays the role of exactly the kind of switch we are looking for. The basis of this suggestion lies in the fact that acetylcholine has different effects on different parts of CA3. Just like the cortical layers, CA3 also has a layered structure. The pyramidal cells of CA3 are located in a layer called *stratum pyramidale* (Latin for a more mundane “pyramidal layer”). Input to these cells are segregated in different layers, depending on their origins. Inputs from neurons in entorhinal cortex, the port of entry to hippocampus, arrive in a layer known as *stratum lacunosum-moleculare*, while the recurrent inputs coming from other CA3 pyramids lie in another layer called *stratum radiatum* (Fig. 5.14). In other words, inputs coming from outside, and recurrent inputs in CA3 are cleanly segregated in separate layers of CA3. To the immense delight of theoretical neuroscientists, who would like the brain to play to the tune of theoretical constructs, it was discovered that acetylcholine has a differential action on the different layers of CA3. It seems to suppress synaptic transmission more in *stratum radiatum* than in *stratum lacunosum-moleculare*. That is, it suppresses the recurrent connections more than the external connections.



**Fig. 5.14** Layers of hippocampal field CA3 that are differentially innervated by cholinergic projections

Thus acetylcholine is acting like a gate-keeper, or a valve if you may, that controls the flow of information in/out of hippocampus, and switches between the dual modes of storage and recall. During storage, when acetylcholine is high in CA3, the recurrent connections are suppressed and external inputs are enabled. Acetylcholine in CA3 also enhances plasticity among CA3 pyramidal neurons. More pieces have now fallen in place. In addition to the general associative network concepts, Hebbian learning, recurrent connections, etc—which we see at work within the hippocampal circuitry, we now also discovered that the important mechanism of switching between storage and recall dynamics is provided by the cholinergic system.

Another question is still unanswered: what triggers the switching mechanism? How does the cholinergic system know when to operate the hippocampal machinery in storage mode and when in recall mode? A beginning of an answer to this question lies in the fact that there is a feedback projection from hippocampus to medial septum. Electrical stimulation of this hippocampal-septal pathway suppresses the activity in medial septum, thereby reducing acetylcholine release into hippocampus. What might the possible function of this feedback branch?

According to Hasselmo and colleagues, this feedback branch has exactly what it takes to complete the memory recall/storage machinery in hippocampus. They suggest that the CA1 module, about which we have not said much in the preceding discussion, compares the output of CA3 with the original entorhinal input. This is possible because CA1 receives direct input from the entorhinal cortex and also projections from CA3. When the entorhinal input matches with the output of CA3, it means the pattern is well-learnt and therefore needs to be merely recalled; when they differ, the pattern is not yet completely learnt/stored, and therefore acetylcholine-dependent

storage mechanism should be triggered. Although this hypothesis is conceptually quite appealing, evidence in support of this schema is still being accumulated.

As a concluding note, we refer to another exciting, almost mysterious, aspect of the role of hippocampus in memory. Note that though hippocampus stores memories, it is only a temporary store, a “scratch pad,” as it is often described, of memory. After a brief sojourn, for a duration that stretches over minutes, hours and days, memory moves on to other longer term stores distributed all over the cortex. We have seen two modes of traffic in the hippocampal circuitry so far—in storage where cortical inputs enter and get stored in hippocampus, and recall, where hippocampal stores are broadcast back to the cortex for recall. If recall consists of hippocampus transmitting its contents back to the cortex, then how does hippocampus write to long-term stores in the cortex? By what mechanism and when does hippocampus achieve this transfer?

## Sleep, Dreams, and Memory

A link between sleep, dreams, and memory has been suggested more than 200 years ago and is a subject of a lot of debate and controversy. Exam-related folklore teaches that a good night’s sleep before the day of the exam aids immensely in a perfect recall during testing times, though the demands of exams often prevent the unfortunate student from taking that option. In an experiment by Robert Stickgold and colleagues, aimed to test the link between memory and sleep, a group of subjects were trained on recognizing a set of visual textures. There was not much improvement in performance when the subjects were tested on the same day 3–12 h later. But performance jumped up dramatically when they were tested the following day after they slept for the night. The improvement continued three consecutive days after the original training. But when the subjects were deprived of sleep on the night following the original training, there was no improvement, not only the following day, for two consecutive days after the original training. Only the third day after, some improvement was observed, though the improvement was much inferior to what was seen when there was no sleep deprivation.

Several experiments of this kind demonstrated the role of sleep in memory consolidation. But more refined experiments revealed the role of specific components of sleep on memory consolidation. Sleep is not a blind plunge into a featureless, dark pit of unconsciousness. It is a structured journey, with its sojourns, its alternations, its cyclical passage through a set of brain states, before the subject wakes to the ordinary world of constrained space and time. Two prominent sleep stages are often described: the slow wave sleep (SWS) characterized by low-frequency waves in the Electroencephalogram (EEG) and the rapid eye movement (REM) sleep characterized by rapid movements of eyes seen under the shut eyelids. The REM sleep is also usually associated with dream state as has been repeatedly verified by sleep research.

Sleep researchers Steffen Gais and Jan Born studied the role of SWS on memory consolidation in human subjects. The subjects were trained on a word matching task in which a series of word pairs were presented sequentially on a computer monitor.

After the presentation, one of the words in each word pair was presented to subjects, in response to which they were expected to recall the other. Memorizing words is an example of declarative memory, a known function of hippocampus. For comparison purposes, the subjects were also trained on a task that involves skill memory or procedural memory. In this task, the subjects learnt to copy simple line diagrams by looking at their mirror images, a skill known as mirror tracing. Neural substrate for procedural memory is known to be different from hippocampus, a brain circuit known as basal ganglia. Similar to the sleep/memory study described above, in this study too subjects showed improved performance after a night's sleep. Specifically, subjects showed enhancement in both forms of memory—declarative and procedural. To verify the specific role of hippocampus in the consolidation process, some subjects were given a drug called physostigmine, a drug which increases the life of acetylcholine in cholinergic synapses. Effectively physostigmine stimulates acetylcholine receptors, with an action that is tantamount to increasing acetylcholine. Physostigmine is found to have selectively blocked performance improvement on declarative memory task (word matching) but not on procedural memory task (mirror tracing). The factors that produce this impairment are clear. Physostigmine effectively increased (the lifetime of) acetylcholine in hippocampal synapses which triggers "storage mode." But what is required for memory consolidation is "recall mode" in which information from the hippocampus is shuttled back to the long-term stores in the cortex. Other evidence that has come from rats is that acetylcholine levels in rat hippocampus are actually low during SWS. Furthermore, declarative memory is impaired, just as in presence of physostigmine, when SWS is reduced significantly.

Thus we may visualize the stage-wise progression, like a baton in a marathon race, of a piece of sensory information that is shuttled around the brain. Sensory information that enters the brain, winds its way through one or more deep brain nuclei before it finds itself in the appropriate primary sensory cortex. Henceforth, the sensory stream is processed by a specific hierarchy of cortical areas specializing in various aspects of that particular sensory modality, before it arrives at the sensory association cortex in the inferior parietal area. This association area combines higher level sensory features extracted from specific sensory modalities and constructs higher level abstract concepts out of the same. The sensory association area also projects to the hippocampus. During waking stage, since acetylcholine levels in the hippocampus are high, information entering the hippocampus is stored in hippocampal regions endowed with richly recurrent connectivity. There is also evidence that only certain kinds of information is marked for hippocampal storage, based on the salience of that information to the individual. Later, in the secrecy and silence of the night, when the brain is in deep sleep (the SWS), information that was lodged in the hippocampus is transmitted to various cortical areas, for long-term storage.

An obvious question that crops up at this juncture is: why two stages in memory? Why cannot there be just a single, massive memory store, operating at longer time scales, and serving as the sole destination of all the information that ever enters the brain? An interesting explanation was offered by David Marr, proposing a need for a dual-stage memory—first short-term memory, followed by a long-term storage. Marr visualized hippocampus as a fast but short-term memory, in which new learning

may overwrite older information. Cortex is offered as a vast memory space, where learning occurs more slowly, but is retained for much longer times, if not forever. Since sensory information impinging on the brain from the external world, flows in rapidly, the slow cortical mnemonic machinery may not be able to successfully capture the influx. Therefore, a fast, limited in capacity, and short-term, memory is needed to rapidly grasp and store the sensory influx. Then, in the leisure and the independence from the external sensory world, offered by the sleep state, the information stored earlier in the hippocampus is slowly passed over to the long-term cortical stores. Thus the known neurobiology of hippocampus and memory dynamics seems to support Marr's speculations about hippocampal function. However, it must be noted, as is the case with a lot else in brain science, the itinerary of memories in the real brain, is a lot more complicated than the neat picture presented in this chapter. A great many mysteries of our memory mechanisms are still unsolved. For example, where exactly are the long-term stores in the cortex? Why does memory consolidation occur in sleep, particularly in deep, dreamless sleep? Are there more memory stages than two? Furthermore, existing knowledge is too vast to be compressed in a single chapter. For example, we have not visited certain types of memory consolidation that occur in dream sleep, technically known as Rapid Eye Movement (REM) sleep. We have not described the brain circuits that are engaged in memorizing motor skills, like riding a bike. Nor have we described memories of emotional events that are processed by another tiny yet important brain structure known as amygdala.

The present chapter is about *temporal* stages of memory, particularly the two memory stages—hippocampus and cortex. We now turn our attention to how information is distributed *spatially* in the brain.

## References

- Briggs, J. P., & David Peat, F. (1984). *Looking glass universe: The emerging science of wholeness*. New York: Simon and Schuster Inc.
- Carmichael, L. (1959). Karl Spencer Lashley, experimental psychologist. *Science*, 129(3360), 1410–1412.
- Gais, S., & Born, J. (2004). Low acetylcholine during slow-wave sleep is critical for declarative memory consolidation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7), 2140–2144.
- Gluck, M. A., & Myers, C. E. (2001). *Gateway to memory: An introduction to neural network modeling of the hippocampus and learning*. Cambridge: MIT Press.
- Halliday, D., Resnick, R., & Walker, J. (2001). *Fundamentals of physics* (6th ed., pp. 866–870). USA: Wiley.
- Hasselmo, M. E. (1999). Neuromodulation: Acetylcholine and memory consolidation. *Trends in Cognitive Sciences*, 3, 351–359.
- Hasselmo, M. E., Schnell, E., & Barkai, E. (1995). Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience*, 15, 5249–5262.
- Hebb, D. (1949). *The organization of behaviour*. USA: Wiley.

- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Sciences of the United States of America*, 79, 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, 3088–3092.
- Kopelman, M. D., & Corn, T. H. (1988). Cholinergic ‘blockade’ as a model for cholinergic depletion. A comparison of the memory deficits with those of Alzheimer-type dementia and the alcoholic Korsakoff syndrome. *Brain*, 111(Pt 5), 1079–1110.
- Lømo, T. (2003). The discovery of long-term potentiation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1432), 617–620.
- Malenka, R. C., & Bear, M. (2004). LTP and LTD: An embarrassment of riches. *Neuron*, 44(1), 5–21.
- Malenka, R. C., & Nicoll, R. A. (1999). Long-term potentiation—A decade of progress? *Science*, 285(5435), 1870–1874.
- Marr, D. (1971). Simple memory: A theory for arhcortex. *Proceedings of the Royal Society, London, B*262(841), 23–81.
- Morris, R. G. M., Davis, S., & Butcher, S. P. (1990). Hippocampal synaptic plasticity and NMDA receptors: A role in information storage? *Philosophical Transactions of the Royal Society of London B*, 329, 187–204.
- Penfield, W. (1952). Memory mechanisms. *AMA Archives of Neurology and Psychiatry*, 67, 178–198.
- Pribram, K. H. (1987). The implicate brain. In B. J. Hiley & F. David Peat (Eds.), *Quantum implications: Essays in honour of David Bohm*. UK: Routledge.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, 20(1), 11–21.
- Solomon, P. R., Groccia-Ellison, M. E., Flynn, D., Mirak, J., Edwards, K. R., Dunehew, A., et al. (1993). Disruption of human eyeblink conditioning after central cholinergic blockade with scopolamine. *Behavioral Neuroscience*, 107(2), 271–279.
- Stickgold, R., Hobson, J. A., Fosse, R., & Fosse, M. (2001). Sleep, learning, and dreams: Off-line memory reprocessing. *Science*, 294, 1052.
- Treves, A., & Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, 2(2), 189–199.
- Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3), 374–391.

# Chapter 6

## Maps, Maps Everywhere



*When one of these flashbacks was reported to me by a conscious patient, I was incredulous. For example, when a mother told me she was suddenly aware, as my electrode touched the cortex, of being in the kitchen listening to the voice of her little boy who was playing outside in the yard.*

—Wilder Penfield.

How does the brain see? This important question turned out to be the theme of a curious cocktail party conversation between V. S. Ramachandran, eminent neurologist, and a young individual uninitiated into the subtleties of brain science. The young chap asked Ramachandran what he did for a living. The neurologist explained that his research interests lay in the study of vision, of how the brain perceives objects. The man seemed to be unimpressed. What is there to be studied in something as simple and spontaneous as seeing? Ramachandran asked how, according to this gentleman, brain saw?

“It’s quite simple,” the youth explained. “Light entering the eye falls on a sheet of light-sensing cells, called the retina, located at the back of the eye. The light-sensing cells, photoreceptors as they are called, convert the light energy into electrical signals. The retina then transmits the electrical signals to the back of the brain via a bundle of nerve fibers and … voila! The brain sees!”

The youth must have noted the dismay on the neurologist’s face to hear such a valorous oversimplification of a deep question.

“Of course, it is not as simple as it sounds. There is a subtlety,” the youth continued to explain as if to quickly repair his reputation. “The image that falls on the retina is an upside-down image of the external world, since light rays crisscross as they pass through the narrow aperture, the pupil, located at the front of the eye. This jumbled image isunjumbled, so to speak, by a complex system of wires that connect the retina to the brain. Therefore, the brain, and thence ‘I’, am able to see an upright image.”

What the young gentleman above has presented is not just an oversimplification of the problem of vision. He seems to suffer from an illusion of understanding, a fairly common malady among those who are interested in brain science. For a large number of people who have spared some casual thought to the question of “how brain sees,” seeing occurs, as if by magic, when the electrically coded signals pertaining to the retina image somehow wind their way into the brain. The obvious philosophical objection to such a simplistic view is that: if brain signals are simply a faithful copy of retinal image, why does seeing not occur in the eye itself? Why do we need the brain to see?

The question of the fate of the image information that journeys beyond the retina onward into the recesses of the brain is more than a century old. Work with experimental animals during the last decades of the nineteenth century, revealed that occipital lobe is critical for seeing because animals with lesions in the occipital lobe fared poorly at tasks that involve vision. In the second decade of the twentieth century, a great wealth of knowledge about the cerebral destinations of optical information was gained, not due to a conscious large-scale research effort in visual neuroscience, but inadvertently, by the study of the unfortunate victims of World War-I.

Gordon Holmes was a British neurologist whose work with WW-I victims became one of the earliest sources of our understanding of visual areas of the brain. Holmes encountered a large number of soldiers who suffered head wounds—in fact 25% of all penetrating wounds. A good number of these cases involved injury to lower occipital area and cerebellum, thanks to the unfortunate design of British helmets that did not adequately protect the lower part of the back of the head. The German helmets were more protective, but since they are heavier, German soldiers did not always wear them, adding to the casualties on their side. The trench warfare, which protected the body but rendered the head more vulnerable, increased suffering on all sides.

Holmes’ work was most probably inspired by the work of Salomon Henschen, a Swedish neurologist and pathologist at the end of nineteenth century. Henschen studied cases of hemianopsia, a kind of blindness in an entire half of the visual field, due to lesions in a part of the cortex, a part that was later identified to be involved in vision. This type of blindness that occurs due to cortical damage must be distinguished from blindness caused by damage to the eyes or the nerve fibers that lead from the eyes to the brain. Henschen correlated the lesions with the type and location of blindness in his patients and noted that damage to areas around occipital pole is associated with blindness, partial or complete. In an attempt to associate the exact location of the lesion with the part of the visual field where the patient is blind, Henschen observed that lesions in lower part of occipital area, typically affected vision in upper visual field. Thus, there seemed to be a spatial correspondence between parts of the visual field and locations on the occipital cortex. These findings prompted Henschen to propose the idea of “cortical retina,” a map of the visual space on the visual cortex. Although the idea of a cortical visual map is on the right track, Henschen was not able to construct the map in detail since the lesions of the brains he studied were often extensive and not sufficiently localized.

Holmes had the advantage of dealing with brains in which small localized wounds were produced by shrapnel causing scotomas, or small, localized blindness in specific areas of the visual field. Shrapnel lost most of its kinetic energy after penetrating the skull and, therefore, made small localized wounds. Holmes was able to correlate the exact location of the wound with the location of the scotoma in the visual field. When the wound was located below (above) the occipital pole, the scotoma was located in the upper (lower) part of the visual field. If the wound was on the right (left), the scotoma was on the left (right). Thus, there was indeed a map of the visual field, and therefore also of the retina, on a part of the occipital cortex known as the primary visual cortex, since it is the first cortical stopover of visual information streaming in from the eyes. Contemporary visual neuroscience offers detailed charts that precisely map points on the visual field to points on the visual cortex. This map of the visual space, and also of the retinal space, onto the visual cortex is known as a retinotopic map.

While discovery of specific modules that are responsible for specific functions marks a significant progress in our understanding of the brain, the realization that information in specific module has a certain orderly spatial organization within the module represents a further step in that understanding. Discovery of maps in the brain is not limited to visual areas of the brain, though the visual maps have inspired search for maps in other sensory modalities.

Wilder Penfield's explorations of the brain using electrical stimulation technique may be considered as some of the earliest investigations into sensory maps. As described in Chap. 5, these stimulation studies were often done to distinguish intact brain tissue from a pathological one—probably a tumor, or an epileptic focus. When Penfield stimulated a cortical region that is right behind (posterior to) an important landmark in the brain, a valley or a sulcus that runs nearly vertically through the middle of the brain, on either sides—a feature known as central sulcus—the patient reported that he/she felt being touched. Further local exploration revealed that it is the part of the cortex that processes touch information coming from the entire body surface, and is known as the somatosensory cortex (somatosensory is simply “body sense”). Stimulation of specific sites in this cortical area produced the feeling that a precise, corresponding body part of the subject has been touched. This correspondence between body surface and the somatosensory cortex seemed to follow a certain “adjacency principle.” Just as in the visual maps, nearby points on the retina are mapped onto nearby points of the visual cortex, stimulation of nearby points on the somatosensory cortex created the feeling of touching nearby points on the body surface. This property of brain maps earns them the title of “topographic brain maps,” suggesting that they are maps of the sensory space onto the brain’s surface. Another curious property of this touch map is that the cortical real estate is disproportionately allocated to its corresponding body surface. For example, the area allotted to the hand area, or to the tongue and surrounding regions, is relatively much larger than the area allotted to the entire trunk. Thus, there is again a map of the entire body surface on the cortical surface (Fig. 6.1).

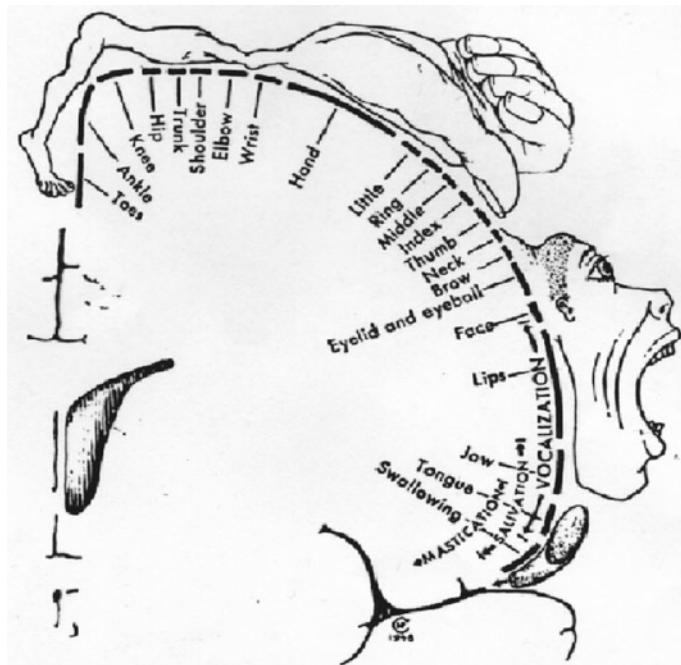
Map structure has been found in brain areas that control motor function too. Some of the pioneering work in this area also came out of the bold surgical studies of Wilder

**Fig. 6.1** A cartoon depiction of the somatosensory cortex. Note that hand and tongue areas are allotted disproportionately large cortical area compared to larger body parts like trunk and lower limbs



Penfield. Stimulation of specific sites in a cortical area a little before (anterior to) the central sulcus produced unintended body movements. When certain sites were stimulated the patient produced rapid, jerking movements of an entire arm or a leg. Stimulation at other locations produced small twitches of one or a few muscles. Further exploration showed that is also a motor map of sorts in this cortical area that controls movements, an area broadly labeled as motor cortex, which was later subdivided into several specialized areas. Existence of motor maps was also anticipated by British neurologist Hughlings Jackson, who arrived at this conclusion from observations of seizure patients. When a seizure strikes, the convulsive movement often begins at a small part of the body, with the tremulous movement gradually spreading over nearby body areas, ultimately involving the entire body. Pondering over this pattern of seizure progression, Jackson visualized a possible development of brain events that lead to seizure: electrical activity arising out of an epileptic focus, and spreading gradually to the nearby areas of cortical surface, like a forest fire, seemed to produce the convulsive pattern. Such pattern of control also implied that nearby points on the motor cortex controlled nearby body parts. Therefore, the motor map, the control center of all bodily movements, is sometimes dubbed the homunculus (literally, a “little man”) which controls the body, the way a puppeteer controls his puppets (Fig. 6.2).

Armed with more sophisticated neuroscientific methods of the last few decades, researchers have unraveled brain maps at a level of precision hitherto unheard of. These rigorous studies have greatly improved on the coarse early maps delineated by Penfield and others. For example, it was found that in the postcentral gyrus, the



**Fig. 6.2** A depiction of the motor map that emerged from Penfield's stimulation studies

general location of the somatosensory map, there is not one but several complete maps of the body surface, each map specializing in one form of tactile information—like vibratory touch or soft touch. Similarly, in the visual cortex, there have been found several visual areas each specializing in specific visual properties, like color, form or motion. The motor maps too were found to be more complex than the simplistic maps constructed out of brain stimulation experiments. Early motor maps were depicted as though they strongly satisfied the “adjacency principle”—nearby brain regions control nearby body parts. But more recent maps showed that only a coarse form of adjacency is displayed by motor maps. Furthermore, maps were found not only on the cortex but in subcortical structures also. For example, the striatum, a nucleus that is a part of an important subcortical circuit known as basal ganglia, has maps of sensory and motor information. Similarly, hippocampus, the short term memory structure visited in the last chapter, has a map of surrounding space. Maps were not exclusive prerogative of human brains. They are found in nonhuman primates, like the apes and monkeys, and lower mammals like bats and rats. If you climbed the ladder of the species further down, you will find maps in bird brains too, dispelling any attitude of condescension toward avian intelligence. And there are maps in the spinal cords of frogs. In summary, topographic maps seem to be a deep fundamental feature of the whole wide neural universe.

But how did these brain maps form? At first glance, the question may not seem more meaningful than asking why the sensory, motor and other areas ended up where they did on the brain surface? Or how the grooves and bumps, and the two hemispheres of the brain formed? A sweeping answer given to a lot of “how’s” in biology is: evolution. Evolution, or molecularly speaking, genes decide every spatial detail and temporal process in the brain, or any other piece of biological matter. Or so it seemed, since the heyday of genetics when the genetic code was cracked and the hunt for genetic basis to a variety of biological phenomena—normal and pathological—began. The first brush with understanding the emergence of brain maps was on genetic lines. Genes, it was believed, control and shape the intricate map structure found in the sensory and motor areas of the cortex and elsewhere in the brain.

But slowly evidence began to accumulate to show that the maps are not cast in stone, genetic or otherwise—they change. The maps are dynamic and seem to adapt to changing conditions of stimulus. A noted, recent example of the dramatic mutability of these maps is the study of somatosensory maps in primates by Michael Merzenich and colleagues. In one of a whole line of studies, Merzenich’s group surgically amputated one of the digits/fingers of an owl monkey and studied the areas of the somatosensory map that respond to the fingers of the same hand. A few months after the amputation, the group found that the maps changed radically compared to what they were before the amputation. Neurons in the map region that earlier responded to the missing finger now responded to adjacent fingers, as though they are gainfully reemployed to serve neighboring regions. Similar data regarding changing maps came from a study of visual cortex and response of visual neurons to changing conditions of visual stimuli. Such findings forced some rethinking of the genetic angle to map formation. Thus, emerged a new line of thinking that viewed map changes as a natural aspect of learning and memory in the brain.

## The Self-organizing Maps

Mathematicians have pondered over the possible mechanisms by which the topographic maps of the brain form. Notable pioneering efforts in this direction include those by Shun-ichi Amari, Stephen Grossberg, Christopher van der Malsburg, David Willshaw and Teuvo Kohonen. They described models of neural networks in which neural responses are shaped by learning mechanisms. Neurons in these models are organized as spatial, two-dimensional layers. Due to learning mechanisms underlying these models, neurons organize themselves such that they collectively form spatial maps of the input space which they respond to. The ability of these networks to organize themselves is described as self-organization. Therefore, Teovo Kohonen, one of the pioneering map researchers called these maps the *self-organizing maps*. These map models have been successfully applied to simulate sensory maps of visual, somatosensory and auditory systems; motor maps of both the motor cortex and spinal cord; sensory and motor maps of subcortical nuclei like the superior colliculus, the

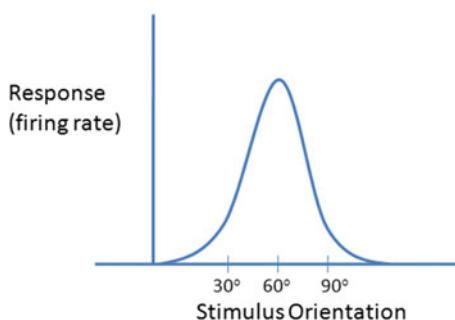
mid-brain structure that controls eye movements, and so on. The self-organizing maps have also been applied to a large variety of engineering problems, since the ability to map abstract types of information neatly on a two-dimensional surface, a sheet of neurons, provided a remarkable means of visualizing data that is essentially high-dimensional and therefore not easy to visualize.

We are not at the moment interested in the engineering applications of the self-organizing maps. But we will discuss, in simple intuitive terms, some of the key ideas involved in these map models, staying clear of the complicated integro-differential equations that the map modelers have toyed with to rigorously explain map formation.

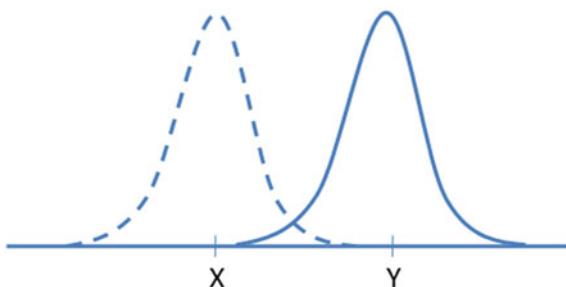
One of the ideas that we must familiarize ourselves with before we dig deeper into map secrets is the existence of neurons with “tuned responses.” Just as a radio can be tuned to receive signals at or around a single frequency, there are neurons that respond to a very special or specific stimulus; its response drops sharply as the stimulus changes and gradually becomes something else. For example, there are neurons in the primary visual cortex that respond whenever a straight bar of a given orientation is flashed in a certain part of the visual field. The response of such neuron often reduces gradually when the orientation of the stimulus is gradually altered (Fig. 6.3). Then there are frequency-sensitive neurons, neurons that respond to a pure tone, in the auditory cortex. There are neurons in the somatosensory cortex that respond to a vibratory stimulus with a specific frequency of vibration, presented at a specific spot of one’s palm of the hand. Similarly, there are neurons in the visual areas of the inferior temporal area that respond to complex visual patterns like faces. An idealization of such face-responding neurons is the “grandmother cell,” a neuron that is thought to respond specifically to the face of one’s grandmother. Although there is no evidence that neurons care to respond to grandmothers, one’s own or another’s, there is curious evidence of neurons that respond to celebrities like Jennifer Aniston!

Thus neurons with tuned responses are real. Their response pattern is depicted abstractly by bell-curves of the kind depicted in Fig. 6.3. The  $x$ -axis represents a parameter of the stimulus, like frequency of an auditory stimulus or orientation of a visual stimulus, while the  $y$ -axis represents the response of the neuron, its firing rate. The response is maximal for a specific magnitude and pattern of the stimulus and trails off in all directions away from the maximum. A natural question that arises

**Fig. 6.3** Schematic of the response of a neuron responding to an oriented bar. Response (firing rate) is the highest at a specific orientation ( $60^\circ$ )



**Fig. 6.4** A neuron initially shows tuned response to a pattern X, which was presented frequently. But later, when a different pattern Y is presented repeatedly, the neuron adjusts its tuning to the new pattern

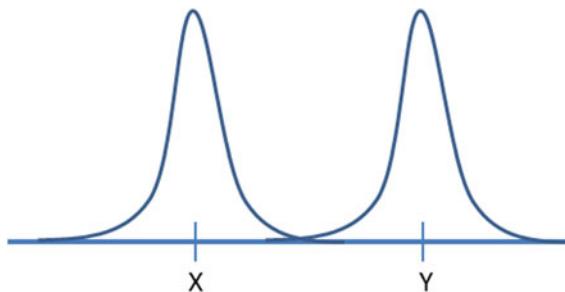


in this context is: how do neurons end up with tuned responses? Although some “tunedness” could arise due to hard-wiring, there is a lot of evidence that tuning can emerge as a result of synaptic plasticity. Specifically, it can be easily argued that plasticity mechanisms like the Hebbian learning could contribute to tuned neural responses.

Consider, for argument’s sake, a pair of neurons A and B connected (from A to B) by a synapse,  $S_{AB}$ , which is of moderate strength. Therefore, assume that normal firing of A initially has only a fifty-fifty chance of firing B, when suitably aided by inputs from other neurons that project to B. But let us assume that, with Hebbian learning in action at the synapse  $S_{AB}$ , even with this moderate level of simultaneous firing of A and B, the synaptic strength gradually increases, to a point when firing of A succeeds in firing B with near certainty. Thus, in this simple situation, we may observe that initially neuron B is not “tuned” to the input from neuron A; but due to Hebbian mechanism acting on the synapse  $S_{AB}$ , B became tuned to A, since it now responds strongly to inputs from A. Since firing level of A is high,  $S_{AB}$  has increased. In case of low firing level of A,  $S_{AB}$  would continue to be low. Thus, in a sense,  $S_{AB}$  evolves to mirror the average activity of A.

Now imagine a similar process occurring over a large number of synapses of neuron B, with different levels of synaptic activity. Synapses that are more active have a chance of getting strengthened. Weaker synapses might even grow weaker as the pre- and postsynaptic activations occur in an increasingly out-of-sync fashion. Thus, the pattern of synaptic strengths of neuron B will increasingly begin to look like the average input pattern to the neuron. Thus, synapses of neuron B will allow B to respond optimally only to certain frequently occurring stimulus pattern X. Thus, B just evolved a tuned response to X. This tuning of a neuron to a stimulus may vary when the stimulus conditions vary. Once B evolves tuned response to stimulus pattern X, let us stimulate B with a new stimulus pattern Y. The tuning of B now changes from X to Y. As B now begins to get tuned to Y, its tuning curve now moves toward the right so as to respond optimally to “Y” (Fig. 6.4).

To understand the mechanisms that underlie map formation, we must add to the idea of tuned responses, the idea of competition among neurons that are evolving tuned responses to specific patterns. Now let us visualize two neurons A and B, both of which receive identical inputs over their synapses (Fig. 6.5). Note that A

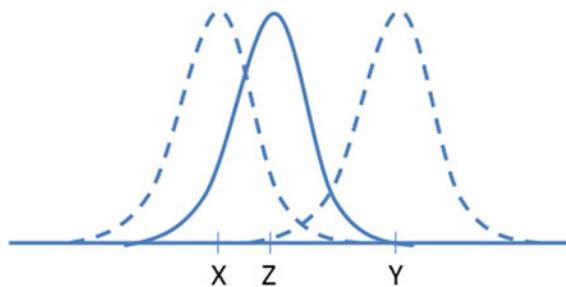


**Fig. 6.5** A schematic showing the tuning curves of two neurons tuned to stimuli X and Y respectively

responds maximally to stimulus “X” and B to “Y.” But now a new pattern “Z,” which lies somewhere between “X” and “Y,” is repeatedly presented. Assume, however, that “Z” is closer to “X” than to “Y.” Now let us introduce competition between neurons A and B, allowing them to vie to respond better to “Z.” Since Z is closer to the tuning curve of A than that of B, “Z” produces slightly stronger response in A than in B. Now imagine that whenever A “succeeds” in responding to “Z,” it sends out a signal to B preventing it from responding to “Z.” The tuning curve of A now moves slightly toward “Z,” and that of B actually moves further away from “Z.” As this process continues, A’s tuning curve moves right on top of “Z,” while B’s tuning curve moves so far off that, in its current state, B hardly responds to “Z” (Fig. 6.6). This kind of competition is known as “winner-takes-all” since the winning neuron continues to win the competition, a process that culminates in completely throwing B out of competition. An example of such competition is the competition among neurons of primary visual cortex to respond to inputs coming from the eyes. Each neuron in a developed visual cortex typically responds to inputs coming from a single eye—right or left, not both, a property known as ocular dominance. But during early development, neurons respond to both eyes. By the pressure of competition, neurons evolve tuned responses and respond to inputs from a single eye.

Having considered tuned responses and competition among neighboring neurons to respond to common stimuli, we are now ready to consider a scheme for map formation. Consider, not just a pair, but a sheet of neurons, resembling the cortical sheet, arranged in a regular grid for mathematical convenience. Initially responses of the map neurons are random, and the map is not ordered, i.e., it does not possess the two important map properties: (1) similar inputs produce activity in nearby neurons in the map (“hand” and “forearm” areas are contiguous in somatosensory map) and (2) more frequently occurring inputs are allocated larger portions of the map (“hand” area is much larger “forearm” area). These changes will emerge as a result of the two mechanisms we discussed above: (1) development of tuned responses and (2) competition among neighboring neurons.

Let us now present stimuli one after another to the map neurons. Each stimulus is presented to all the neurons. Competition is set up over the entire map to decide

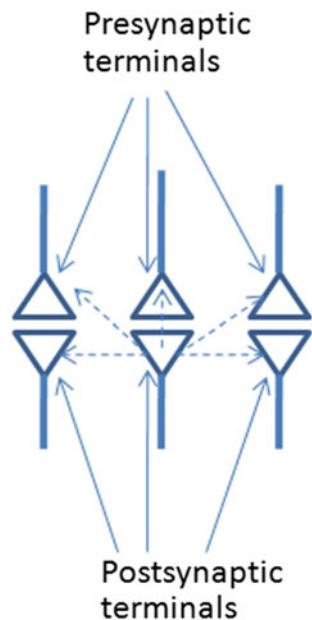


**Fig. 6.6** Schematic depicting competition between two neurons to respond to a stimulus. The dashed left tuning curve, tuned to X, is the initial response of neuron A. The dashed right tuning curve, tuned to Y, is the initial response of neuron B. After repeated presentation of pattern Z, neuron A adjust its tuning to respond to Z

the winning neuron. The neuron which produces the highest response to the stimulus is the winner, whose tuning curve is slightly shifted toward the stimulus. The key aspect of map algorithm is: in addition to the winner, the tuning curves of the neurons in the neighborhood are also adjusted. This neighborhood is typically divided into a “central” portion, the portion that forms the immediate vicinity of the winner, and the “surround,” the portion that lies beyond the central portion, but encloses the winner. The tuning curves of the central neurons are shifted toward the input, and those of the neurons in the surround are shifted away from the input. Thus, competition is not between pairs of neurons, but groups—the central ones against those of the surround. Note that since neurons in the “center” always move together toward the stimulus, they actually cooperate. Due to such cooperation, nearby neurons evolve similar responses, which is one of the key properties of the map. The competition between the center and surround neurons ensures that there are always outlying neurons that do not move toward the present stimulus. They will then be available to respond perhaps to another stimulus. Furthermore, if there are a large number of stimuli of a certain kind, they fire a group of neurons repeatedly, which in turn drag their neighboring neurons toward the same stimuli. Thus, types of stimuli which are more numerous, or parts of the input space that are more populated, are allocated larger areas of the map. Thus, the simple combination of Hebbian learning, local competition/cooperation among neurons offers the machinery necessary to form maps that can emulate the ubiquitous cortical maps in the brain.

We have already reviewed the biological evidence of Hebbian learning by linking it to Long-Term Potentiation (LTP). But is there evidence for local competition and cooperation among neurons? Basically, the above map mechanism requires a signal that allows synaptic plasticity in one neuron to influence plasticity in nearby neurons. It has been proposed that nitric oxide (NO), a gas that is released by active neurons, can play the role of such a messenger. Nitric Oxide is known to be a sort of a “hunger signal” emitted by active neurons which results in dilation of nearby microvessels. Vascular dilation results in the release of oxygen and glucose necessary for fueling neural activity. In addition, NO is also thought to be one of the factors that influence

**Fig. 6.7** NO (dashed arrows) is released from the postsynaptic terminal of an active synapse. It diffuses to the presynaptic terminal of the same synapse, and also to nearby synapses inducing LTP



LTP. NO is a gas that diffuses rapidly in the extracellular space surrounding neurons. When released from a neuron it can act on the same neuron's synapses and enhance LTP in them. Since NO spreads to nearby regions, it can also enhance LTP in the synapses of nearby neurons, thereby acting as an agent of cooperation required for map formation (Fig. 6.7). NO is destroyed within seconds by oxidation. Therefore, it has a limited domain of influence that depends on its diffusion constant (how far it can spread in unit time) and its lifetime. In spite of its elegance, the role of NO in map formation is only a hypothesis, though a very plausible one, and awaits more direct experimental verification.

Above, we presented an intuitive picture (without the gore of mathematical equations) of the self-organizing mechanisms underlying map formation. The self-organizing map models have been successfully applied to a variety of real-world domains ranging from monitoring the health of electrical transformers to classifying text documents. We now show how the map models have been applied to explain formation of a variety of brain maps.

## Mapping the Bat's Brain

Animals find their way about in the wilderness using the sights, sounds, and smells they receive from the surroundings. Some animals go beyond passive reception of environmental stimuli. They actively emit sounds, and by listening to the sounds

that bounce off the surrounding objects and return, they estimate the distance, relative velocity, and orientation of their targets—a process known as *echolocation*, or locating objects by echoes. Whales and dolphins echolocate in the sea using sonar signals, as do modern submarines. Some cave-dwelling birds resort to this technique to find their way in the pitch darkness of the caves. A creature that is not strictly a bird, but a flying mammal that is known for its uncanny ability to echolocate is the bat. Although bats possess eyes, they navigate using echolocation, a fact that was first established in the eighteenth century by Italian physicist Lazzaro Spallanzani by a series of delicate experiments.

The principle underlying echolocation is similar that of the radar where the distance ( $d$ ) to the target is calculated using the time delay ( $\Delta T$ ) between the emitted signal and the received signal, and the speed of sound ( $c$ ), using the following formula:

$$d = \frac{\Delta T}{2c}$$

The intensity of the returning sound signal, which is weaker than the emitted sound, also provides additional information about the target's distance. The return signal also conveys information about the target's azimuth, the angle that tells you whether the target is to the right, or left or straight ahead. This information is coded in the delay between the times when the return signal reaches the two ears of the animal. When the target is to the right, for example, the return signal reaches the right ear first, than the left ear. Furthermore, the return signal arriving at the right ear is stronger, since the signal loses some of its energy as it passes through the body of the animal's head to reach the left ear, providing additional information to estimate azimuth. There is indeed evidence for existence of neurons that are sensitive to these time delays in bat's brain.

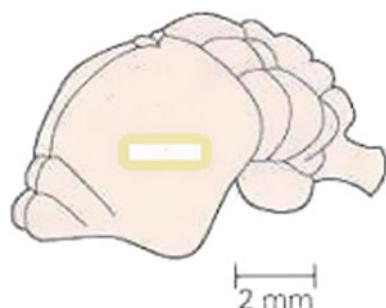
Another piece of information that bats derive from the echoes of their shouts is the relative velocity with respect to the targets. Relative velocity information might come particularly handy when the animal is on a hot pursuit of a prey. Relative velocity estimation depends on a special auditory effect known as the Doppler effect, named after its discoverer, Austrian physicist, Christian Doppler. Doppler effect relates the pitch of a sound coming from a moving source with the velocity of the source. A familiar context in which Doppler effect can be experienced is a moving, hooting train. As the train approaches, a bystander finds the pitch or the shrillness of the sound of the horn gradually increasing; once the train passes the bystander and begins to move away, the pitch drops rapidly. Bat uses this principle to find the relative velocity of its target. If the echo that returns from the target has a smaller pitch, it means the target is receding from the animal. The pitch is greater in a situation when the target approaches. Thus, by comparing the pitch of the sound emitted, with that of the echo, bats are able to find the speed at which their targets are moving from themselves.

To perform such complex auditory operations, the bat's brain must naturally possess well-developed auditory machinery. The auditory cortex of a particular kind of bat known as *Pteronotus parnellii rubiginosus* has been studied extensively by neu-

robiologist Nobuo Suga and his colleagues. The *Pteronotus* bat is a native of Panama and is known to be able to resolve relative velocities up 3 cm/s. Its sensitive auditory machinery enables it to even detect beating of insects which form its staple diet. The sounds that this bat emits are not humanly audible. The bat emits sound pulses that are 30 ms long in duration, at a frequency of about 61 kHz which is in ultrasound range. Frequencies of the echoes are greater than or less than 61 kHz, depending on whether the target is receding or approaching, but they are in the whereabouts of 61 kHz. For the smallest relative velocities that a bat can resolve, it was estimated that the smallest frequency changes that a bat can resolve are as low as 0.02%.

Delicate experiments on the *Pteronotus* bat's brain revealed an amazing organization in the auditory cortex that enables the animal to resolve such minute changes in frequencies. Actually, neurons in the auditory cortex of the bat can respond to wide range of frequencies—from 20 to 100 kHz. But this high resolution of 0.02% is not spread uniformly over this entire range. The high resolution is concentrated around 61 kHz which happens to be the frequency of the sounds emitted by the bat. Figure 6.8 depicts a simplified picture of the bat's brain, with the rectangular strip indicating the primary auditory cortex. Neurons in the primary auditory cortex typically exhibit tuned responses to single frequencies. A careful exploration of this cortical region with electrodes revealed that a large stretch of the rectangular region is devoted to processing sounds in a narrow band around 61 kHz. The smaller flanks are committed to larger bands of frequencies on either side of 61 kHz. The purpose of such disproportionate organization, the likes of which are not found in higher mammals like cat, for example, is not difficult to guess. Unlike a cat, which processes a wide spectrum of sounds received from the environment, a bat depends crucially on a narrow band of sounds around the frequency of its emitted sounds. Therefore, it needs more detailed processing power in a narrow band around 61 kHz, which explains the disproportionate allocation of cortical real estate. Similar disproportionate allocation is seen in other brain maps too. In the somatosensory cortex, the area corresponding to the hand and fingers is nearly as large as the area corresponding to the trunk. Similarly in the retinal map of the primary visual cortex, a small part of the retina known as fovea, is mapped on to a disproportionately large part of the cortex, because of the high resolution of visual processing in the fovea.

**Fig. 6.8** A simplified depiction of a bat's brain. The rectangular strip shows the approximate location of primary auditory cortex



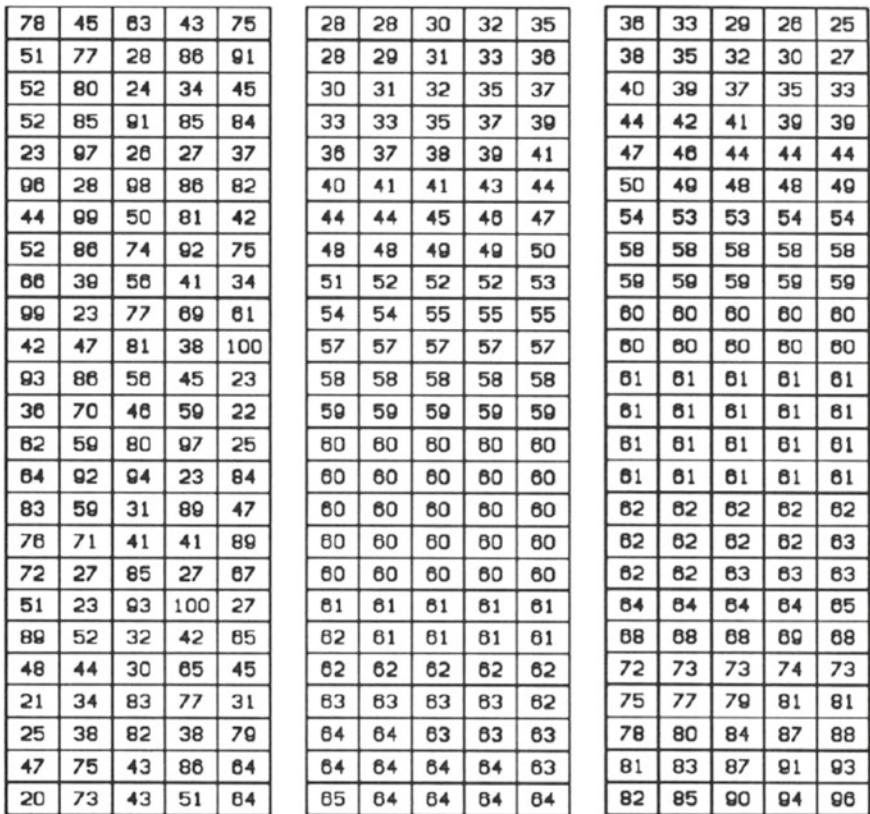
The self-organizing map mechanisms described in the previous section have been successfully applied to explain the organization of the bat's auditory cortex just described. Ritter and Schulzen developed a self-organizing map model containing a rectangular array of neurons with 125 rows and 25 columns. Each neuron has tuned response to a frequency. In the beginning, the neurons are randomly tuned, and there is no pattern in the spatial distribution of the tuning frequencies (Fig. 6.9a). Now single frequencies are presented to the map and the tuning frequencies are adjusted according to map mechanisms described in the previous section. After about 500 learning iterations the map evolved in such a way that there is a continuous map of frequencies along the length of the map (Fig. 6.9b). Another feature of the map at this stage is that nearby neurons in the map respond to similar frequencies. What the map does not have yet is a disproportionate allocation of map area to frequencies around 61 kHz. But on further training of the map, after about 5000 iterations, the map evolved that disproportionate structure. Nearly a half of the map lengthwise is now allotted to frequencies from 59 to 62 kHz, though the map has neurons that respond to a much wider range (30–95 kHz) (Fig. 6.9c). The above map model is strongly indicative that self-organizing mechanisms of the type described in previous section shape the auditory map of the *Pteronotus* bat.

Another system in which the self-organizing map simulations explained both normal maps, and map reorganization under pathological conditions, is the model of somatosensory maps in monkeys.

## Dynamic Reorganization in Somatotopic Maps

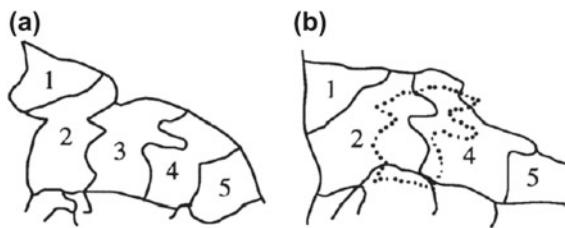
The somatosensory cortex has a map of the entire body surface—hence called a somatotopic (soma=body, topos=space) map, in which nearby spots on the body surface are mapped, for the most part, onto nearby spots on the map. However, this map is not drawn to scale and, like the auditory maps in bats' brains, there is a disproportionate allocation of cortical area to various regions on the body surface. In addition, the map is not always continuous, adjacency-preserving; there are “fractures” in the map where the continuity rule is violated. Also, a simple topological argument suffices to show that a fully continuous map is not possible since the body surface is a closed surface, barring a few “singularities” due to presence of orifices like mouth or nostrils, while the somatosensory map is an open region in the cortex. The fractures seem to arise because the constraints that shape map formation are more involved than simple adjacency preservation.

Disproportionate allocation of cortical map surface to various parts of body surface seems to depend on the amount of information that flows from the body surface. The hand/finger area has a larger map area than the trunk, for example, since we use the rich, touch information that flows from our fingers to probe the world and make a tactile sense of it. Another map region that looks abnormally large, is the tongue and mouth area, expressing the importance of these parts to produce speech, perhaps the most important and prized aspect of human behavior in our increasingly



**Fig. 6.9** Results from a self-organizing map model of auditory cortex of bat. The model consists of a  $125 \times 25$  array of neurons. The numbers in each cell represent the frequency to which the cell is tuned. **a** Left column, denotes the map in the initial stages, **b** middle column, denotes the map after 500 iterations and **c** right column, denotes the map after 5000 iterations (from Martinetz et al. 1988)

extroverted and voluble culture. If the map features depend on something dynamic-like information that flows in via a portion of the body surface, and on some genetic preference to a specific region, it follows that manipulating the information flows from the body surface might alter the map structure. If you sit with your hands tied, one may ask, for days and weeks, do your hand/finger regions shrink? Or if you indulge in extensive body tattooing for long months, does your trunk map expand? It might be inconvenient to perform and test the results of these bold experiments, but analogous results have been observed in less dramatic situations. Experiments have revealed that frequent stimulation of certain spots of body surface led to expansion in the corresponding map regions (Jenkins et al. 1990). Similarly, if the tactile information from certain regions of body surface is cutoff, neurons that would respond



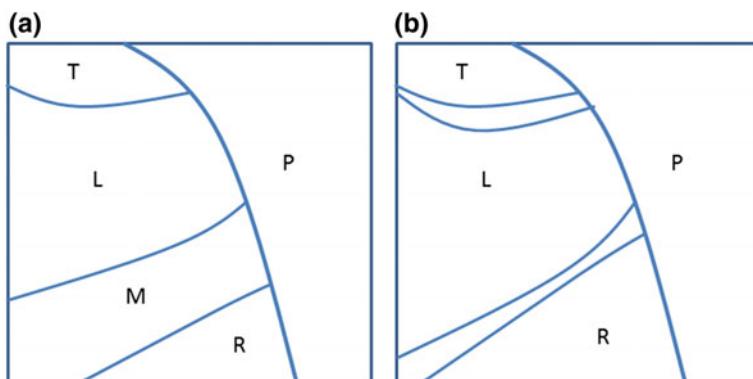
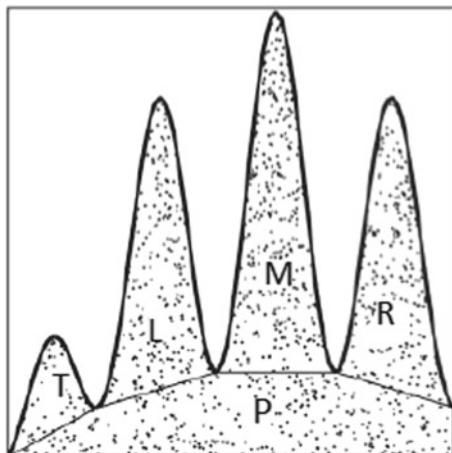
**Fig. 6.10** Adaptation of the somatosensory map corresponding to the five fingers (1, 2, 3, 4, and 5). **a** Cortical regions that respond to the five fingers. **b** When finger 3 is amputated, regions 2 and 4 expanded into what was earlier region 3 indicated with dotted lines (after Fox 1984)

to those regions were found to evolve to respond to inputs from other body regions, near and far.

Jeff Kaas and colleagues studied the dynamic nature of somatotopic map in an adult ape (Kaas et al. 1983). The five fingers of one hand of the animal were mapped. Figure 6.10a shows an idealization of the map, which reveals a nearly linear ordering of the five regions corresponding to the five fingers. Now the middle finger of the animal was amputated and the resulting changes in the map were observed over a long period. Several weeks after the amputation, the region of the map that earlier responded to middle finger started responding to the adjacent fingers—index and ring fingers. With middle finger missing, the region numbered 3 in Fig. 6.10a had nothing to respond to. Gradual rewiring of the inputs to these neurons takes place such that they are gainfully reemployed and begin to respond to neighboring fingers. As a result, regions 2 and 4 now expand and encroach into what was region 3 earlier (Fig. 6.10b).

Such dynamic reorganization of somatotopic map has been modeled by Ritter, Martinez, and Schulten using self-organizing map models. The model is not meant to be biologically detailed; it only aims to capture the essential mechanisms underlying map formation, which it successfully did. The map model consists of a  $30 \times 30$  array of neurons. The inputs are simply points from the image of a “hand,” a two-dimensional area, parts of which designated as the five fingers (see Fig. 6.11). Initially, neurons in the map have tuned responses indeed, but there is no ordering. Nearby neurons do not respond to the same finger, as one would expect, nor to nearby fingers. Responses have random organization initially. However, after about 500 iterations, a coarse map can be seen to have formed. Contiguous stretches of neurons in the map can now be seen to respond to the same finger. However, there are some neurons at this stage that do not respond to any point on the hand surface. On more extensive training, after about 20,000 iterations, a well-ordered map can be observed (Fig. 6.12a). The top right part of the map has neurons that respond to the palm. Below the palm region, there are four nearly parallel strips which correspond to the four fingers. Nearby neurons now respond to the same or adjacent finger quite consistently. Now to simulate the amputation experiment, inputs from the middle finger are omitted, and the map training is continued. After another 50,000 iterations of training, the

**Fig. 6.11** Schematic of the two-dimensional “hand” consisting of five fingers—Thumb (T), Left finger (L), Middle finger (M), Right finger (R), and Palm (P)—used for training the self-organizing map (Redrawn based on Ritter et al. 1992)



**Fig. 6.12** **a** A self-organizing map model of the five fingers formed after 20,000 iterations (Redrawn based on Ritter et al. 1992). **b** Map reorganization after the middle finger is “amputated.” The L and R regions of Fig. 6.11 encroached into what was M region. Two thin strips in the top-left and bottom-middle parts of the map denote neurons that do not respond to any finger (Redrawn based on Ritter et al. 1992)

regions corresponding to the left and right fingers may be seen to encroach into what was earlier the “middle finger region” (Fig. 6.12b).

Thus simple mechanisms of self-organization may be shown to explain dynamic reorganization of the somatotopic maps. But the scope of these mechanisms does not stop at explaining map formation. Concepts related to map reorganization can also explain—and also suggest a cure for—a perplexing clinical phenomenon known as the phantom limb.

## Where Exactly Is My Hand?

What we call the “self”—ourselves—is said to have two parts: the “inner” self and “outer” self. The “inner” self—consists of thoughts, feelings, dreams, etc. The “outer” self—consists of our body. Our inner self of thoughts and feelings is a changing entity all the time. The outer self, the body, has a well-defined shape, a clear outline. The body also does not change very rapidly. This constant sensation of our body is known as the “body image.” It is thanks to this that we feel we are separate from others and other objects in the world.

There are situations when this body image is mysteriously and painfully distorted. When people lose limbs in an accident, or war, or due to surgery, almost always they are left with a lingering sensation of the missing limb. The person feels that the missing arm can still “open doors,” “wave goodbye” to friends, “twiddle on the table,” etc., while in fact there is no arm at all! This living sensation of a nonexistent arm is known as, in medical terminology, the “phantom limb.”

One of the earliest accounts of phantom limbs comes from the life of Lord Nelson, the British naval officer who defeated Napoleon. Nelson lost an arm in war and had phantom arm sensation. This lingering sensation of a missing body part, he believed, “is the proof of existence of soul.”

During the era of American Civil war, Philadelphia-based physician Silas Weir Mitchell coined the term “phantom limb.” In those days, when there were no antibiotics, it was a common practice to amputate infected limbs of wounded soldiers. Most of these amputees had reported phantom limb sensation. The moving, acting, nonexistent limb is not only a source of irritation and embarrassment, it often goes with excruciating pain, and hence poses a serious clinical problem.

There were several theories of the “phantom limb.” One is called “the theory of wishful thinking.” It says that since the amputee badly misses the arm he feels that it is still there. Another theory said that the cut nerve endings in the missing arm turn into “neuromas” which continue to send signals to the brain, which mistakes them as coming from the arm itself. By the same logic, cutting the nerve endings one more time and clearing the neuromas should be a valid treatment for phantom limbs. However, this did not work. Further amputation going all the way to the shoulder, and sometimes higher up to sectioning the spinal cord did not solve the problem. The pain continued.

How do we treat this pain coming from a nonexistent limb? First of all, what or where is this phantom limb? San Diego-based neuroscientist, V. S. Ramachandran, achieved remarkable breakthroughs in understanding this problem. Work on touch sensation in monkeys, performed by Pons, a scientist at National Institute of Health, USA, gave Ramachandran important clues.

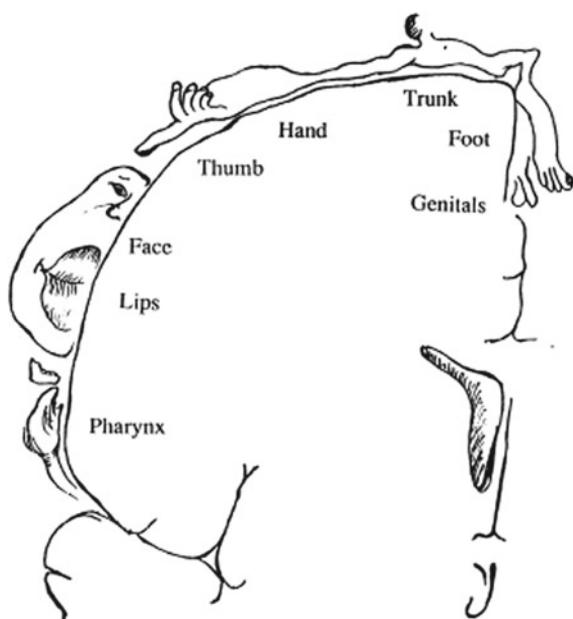
Pons studied monkeys in whom nerves carrying touch information from the hands are cut. Eleven years after this surgery when the hands are stimulated (touched) the corresponding brain area (which would have responded in a normal monkey) failed to respond, which is not surprising since the hand nerves were cut. But what is surprising is that the same area got activated when the monkey’s face is stroked!

V. S. Ramachandran began an in-depth study of the above findings. What has touching the face got to do with hands? The two parts are not even close. At least not in the world where we see them. Though they are not physically close but they are next to each other in the somatotopic map. In Fig. 6.13, you will notice that the portion of the brain map responding to hand is next to the portion responding to the face. Since the “hand” neurons are not responding anymore (no news coming from the hand), the neighboring “face” region expanded and encroached the “hand” region. Stroking the face now activated neurons in erstwhile “hand” region! That simple and brilliant insight opened the doors to many of the mysteries of phantom limb.

Ramachandran wanted to test his insights on his phantom limb patients. One of the first patients was Tom, a 17 year old, who lost his arm in a motor accident. His phantom arm could “wiggle,” “reach out,” and “grasp.” With this patient, Ramachandran did a simple experiment—the classic “Q-tip” experiment. The doctor stimulated specific points on Tom’s (who was blindfolded) body surface with a cotton swab and ask the patient where he is being touched. When Tom’s cheek was touched, Tom confirmed that the cheek was touched but he also said that the thumb of his phantom arm was touched! Similarly, when his upper lip was touched, he said that his phantom little finger was also touched (apart from his upper lip). These results confirmed what Pons observed in the case of monkeys—the “face” region in the brain map has now taken over the “arm” region.

But that’s not just it. On probing Tom further with the Q-tip, the neurologist found that when certain parts of the shoulder were stimulated, the subject reported

**Fig. 6.13** The map of the body surface onto the brain surface



that he felt that his missing hand was being touched. A second entire map of the hand was found on the shoulder on the same side as the missing hand. This finding is not too surprising if we recall from the earlier case of amputated middle finger in the monkey, that the map areas corresponding to the two neighboring fingers of the missing finger took over the map region that had corresponded to the missing finger. In the somatotopic map, the shoulder and the face areas may be found on either side of the hand (Fig. 6.13). Therefore, the encroachment occurred from both directions resulting in phantom sensations in two other parts of the body.

Thus simple rules of map formation are found to be capable of explaining a range of oddities associated with phantom limb experience. But perhaps the greatest contribution of the self-organizing map mechanisms to the phantom limb is to the problem of phantom limb pain. Phantom limb patients often experience their missing limb “doing things” like turning a doorknob, or waving to someone, or holding a coffee cup, or reach for a phone. All this at the best may be regarded as a source of annoyance to the patient. But in some cases, the arm is experienced as being permanently “frozen.” The patient has no control over it. (This particularly happened when the patient had his/her arm held in a sling or in cast before it was finally amputated.) Even this could not have been a problem but for the disturbing fact that the paralytic phantom limb can sometimes be a source of great pain. Movements, or the lack thereof, in a missing limb may themselves sound mysterious, if not absurd. Pain in a nonexistent limb may seem ludicrous but must be taken seriously since it is a fact.

To understand the brain mechanisms that might have precipitated this frozen condition of the phantom arm, we must consider the neural machinery that constantly creates and shapes our body image. Body image consists of our moment-to-moment experience of the shape of our body, of the space it occupies. When we are sitting we know that our knees and back are bent in certain ways, and our weight is bearing down on the seat of the chair. We know this certainly because we can see ourselves in this position, but even with our eyes shut, it is undeniable that we can *feel* ourselves in the sitting position. This feeling arises because of the sensory information—technically called proprioceptive sense or position sense—that comes from our muscles and joints. Muscles have sensory apparatus known as the muscle spindles which convey information about muscle length and changes in the length through time. Tendons, which connect muscles to bones, have receptors that convey information about muscle tension. Joints also have machinery that conveys information about joint angles. In addition, our skin has receptors that transmit events of contact with other objects rubbing on itself. Brain integrates these various sources of proprioceptive information, combines with visual feedback to check for consistency, and constantly creates and upholds a subjective image of the body.

The body image is present not just when the body is in a static position but is rapidly updated and maintained even when it is in motion. Let us consider, for example, how our body image is updated when we lift an arm. Willed movement of the arm, or that of any body part for that matter, is initiated and controlled by a command from the highest center of motor control—the motor cortex. Commands that originate from here trickle down to the spinal cord and proceed toward the

muscles, activating them and producing movement. Movements in the arm produce proprioceptive information which is fed back to two important parts of the brain—the parietal cortex and cerebellum. Interestingly, when the motor cortex generates a motor command, it sends copies of the command exactly to the very same regions which receive proprioceptive information—parietal cortex and cerebellum. Motor command (the copy) received by the parietal cortex and cerebellum informs them of the *expected* changes in the arm position due to execution of the motor command. Proprioceptive feedback received by these two regions informs them about the *actual* changes in the arm position due to the motor action. The difference between the actual and expected changes in the arm position are computed in these two regions. If the difference is zero, the sensory feedback is consistent with the motor action. If the actual is different from the expected, the motor command is subtly corrected, arm position is readjusted, until the difference is zero, and the arm is installed exactly at the desired spot in space.

Let us consider how this entire signaling system works in a phantom limb patient. The motor cortex sends command to the missing arm to move. Since it is nonexistent, and therefore cannot move, the proprioceptive feedback registers no signal. In order to cope with this mismatch, the motor cortex sends a stronger command to move, which is again met with the same null result. As this futile experiment to move is conducted for a while, the motor to sensory feedback loop seems to get broken since the brain realizes that the motor command is ineffectual in moving the arm. This realization seems to result in what is termed a “learnt paralysis.” In some cases, the early attempts to strengthen the motor commands in order to move the arm seems to cause the experience of a “clenched fist” which is associated with pain. The extreme activation from the motor cortex seems to place the phantom limb in this extreme, uncomfortable state. A possible way in which this difficulty could have been corrected and regulated is by introducing an appropriate sensory feedback from the hand—which is impossible. Thus, the phantom hand remains stuck in the “frozen” position, with unrelenting paradoxical pain as a disturbing side effect.

Ramachandran invented an ingenious way of reinserting the missing sensory feedback, thereby closing the loop. Sophisticated ways of introducing such sensory feedback is to fit an artificial arm, or electrically stimulate the nerve fibers that were earlier carrying signals from the missing limb, or use a virtual reality system that can substitute the missing arm with an illusion of it. But all these methods are complicated, cumbersome and expensive. Ramachandran suggests a disarmingly simple substitute for these difficult approaches. This new method requires nothing more sophisticated than a sufficiently large cardboard box and a mirror. This “mirror-box” setup consists of a large box with two holes through which the subject can insert his/her hands (Fig. 6.14). In case of a phantom limb patient, one of the hands is going to be the stump of the phantom hand. The top of the box is open. A mirror is placed vertically, dividing the box into two partitions. The mirror will face the intact hand. The subject will adjust his/her viewing position such that he can see the image of his intact hand in the mirror in the box. From that vantage point, the mirror image of the intact hand will be spatially superimposed on the expected position of the phantom hand. Now the subject is asked to make movements of the intact hand and observe the movements of the image. The image has the uncanny appearance

**Fig. 6.14** Mirror-box arrangement used to treat phantom limb pain



of a moving phantom limb. It is as though the phantom that had been invisible, and immobile all along had suddenly sprung to life and taken on a form. The patient is also asked to visualize that his/her phantom hand is moving in concordance with the mirror image of the intact hand. Through such a cunningly elegant setup the missing sensory feedback from the phantom limb is restored. Subjects who practiced with the mirror-box often reported that they regained movement in their phantom limb. In some cases, the phantom limb was completely exorcized! The phantom limb that troubled its owner with its obstinate immobility simply disappeared. Note that in the mirror-box setup, the sensory feedback that is restored is purely visual; the proprioceptive feedback from the phantom limb continues to be absent. Perhaps in such cases, the conflict between the visual feedback and proprioceptive feedback convinces the brain that the best way to interpret the conflicting situation is to imagine that the phantom limb is gone!

The above “treatment” of problems related to phantom limb is striking by its sheer elegance. It is an excellent example of how a sound understanding of maps, and mechanisms of map formation that support the existence of the phantom limb, can be turned upon itself and effectively used to dissolve the problem. It is ironical—and certainly demystifying—that a phenomenon that was thought to be solid evidence for the existence of the immaterial soul, can be tackled so effectively using a piece mirror and a cardboard box.

Brain maps of touch are straightforward to understand, so are nearly linear maps of frequency in auditory cortices. There are more complex and subtle maps of visual information scattered over the multitude of visual areas of the occipital and parietal areas. But these too can be understood as a comfortable extension of the other kinds of sensory maps we have already encountered. But even more interesting are maps of abstract entities like words, of nouns and pronouns, verbs, adverbs—the whole gamut. Like the sensory maps, words are also mapped onto cortical surface with an intricate logic of their own. These word maps too can be explained with a high success, an example of which we will visit now.

## Mapping the Parts of Speech

Words may be thought of as fragile vehicles for transmitting experience. They may be regarded as abstractions, invariances constructed out of our sensory-motor experience of the world. The sensorimotor bases of words (and perhaps of all language) can be perceived easily if we consider the name of simple object like a “shoe.” The word “shoe” is a token of our entire experience of that object—its appearance, its feel when you walk in it, its squeaking or tapping sounds, its odors pleasant and otherwise. There are other words, like “liberty” for example, certainly more removed from the world of name and form, but traceable perhaps, by extended reasoning, to the sensory-motor experiences that life brings our way.

Although the two hemispheres of the brain are structurally quite similar, language function is predominantly confined to the left side of the brain in over 90% of right-handed people. In about 60–70% of left-handers too language is processed by the left brain. This left-sidedness of language processing in the brain came to limelight for the first time through the studies of French neurologist Paul Broca. Broca studied aphasias, or language disorders, particularly a type of disorder—later named after him as Broca’s aphasia—in which the patient could understand language, could utter single words, or hum tunes, but could not construct complete, grammatically correct sentences. Postmortem studies revealed that Broca’s patients had invariably lesions in the left side of the brain, particularly in the left frontal area, a portion close to the central sulcus, a chasm that divides each hemisphere partially, vertically. This led Broca to announce: “We speak with the left hemisphere” (“nous parlons avec l’hemisphere gauche”).

An opposite difficulty is seen in patients with Wernicke’s aphasia, a form of aphasia first studied by Carl Wernicke a German neuropathologist. Patients suffering from Wernicke’s aphasia speak freely, and may even sound normal, using grammatically correct sentence construction. But the content of their speech is unintelligible, incoherent and flawed. They use random or invented words or substitute wrong words (like “television” for “telephone”) in their sentences. These patients had lesion in a portion of the posterior cortex named the Wernicke’s area, placed strategically among the three important sensory areas—visual, auditory and somatosensory areas. Whereas Broca’s area is responsible for language expression, the Wernicke’s area is involved in language understanding. Furthermore, the fact that the Wernicke’s area is located close to the sensory areas, with strong connections to auditory and visual areas, is an indication of the sensorial roots of language.

Close to the Wernicke’s area, in the temporal lobe and in the inferior regions of the occipital lobe, along what is known as the occipitotemporal axis, there are regions crucial to word-formation, the ability to choose the right word to express a given concept. Eminent neurologist Antonio Damasio describes two of his patients who had special difficulties in accessing words. The patients retained understanding of normal concepts, of common day-to-day objects, like vehicles, animals and buildings, and their functions. But they had difficulty in finding words for objects that are otherwise quite familiar with. For example, when shown a picture of a raccoon,

A. N. will say something like: “I know what it is – it is a nasty animal. It will come and rummage your backyard... I know it, but I cannot say its name.” Thus, these patients had particular difficulty with common nouns, like names of fruits, vegetables and animals. Interestingly they had lesser difficulty with names of tools, perhaps because tools are associated with actions, and action-words like verbs, are processed elsewhere.

Damasio’s patients also had difficulty with proper nouns. When shown a picture of Marilyn Monroe, A. N. said: “Don’t know her name but I know who she is; I saw her movies...” Thus it is clear that these patients are able to retrieve and recognize the ideas and concepts associated with an object, but simply cannot get themselves to name it. Interestingly, the patients did not have much difficulty with verbs, and other parts of speech like prepositions, conjunctions, and pronouns. Their sentences were grammatically sound.

This difficulty in accessing noun forms seems to be linked to the brain area that is damaged—the occipitotemporal axis. Structures necessary for accessing nouns seem to be distributed in graded fashion along the occipitotemporal axis—like a map. More general concepts (associated with pronouns) seem to be mapped onto the rear end, in the posterior left temporal areas. More specific concepts, like the proper nouns, are mapped on to the front, close to the left temporal pole. Patients with lesions in the left temporal pole, therefore, exhibit deficits in accessing proper nouns, but not in common nouns.

The map of nouns—from general to specific—along the length of the occipitotemporal axis is perhaps not unlike the nearly linear map of frequencies on the auditory cortex of bats. But a linear map of frequencies is quite conceivable since frequency is a simple scalar quantity. A frequency is a single number; a noun is not. A noun as an abstract category of words and the noun-map in the brain seems to have a layout of the multiple levels of abstraction of these abstract categories—different types of nouns. How do such maps form? In the absence of an immediate, easy answer, do we hasten to resort to genetics? Or do the self-organizing map mechanisms show us the way in this case too?

Ritter and Kohonen developed a self-organizing map of word forms, known as a semantic map and observed that words that correspond to the same parts of speech or a lexical class (nouns, verbs, adverbs, etc.) cluster together naturally in the map. We obviously know the lexical class of a word when the word is familiar, the way, for example, we know that a mango is a noun. But we can often guess, quite rightly, the lexical class of words that we have never encountered. Language expert Steve Pinker mentions a study performed in 1958 by psychologist Jean Gleason with small children. Four to 7-year-old children were given a simple test, known as the wug test, in which the children were shown a picture of what seemed like a quaint, cartoon bird with the sentence “This is a wug” written underneath. Below this figure, there were two more pictures of the same delightful creature, with the sentence “These are two \_\_\_\_.” The children were required to fill the blank. It turns out that 99% of the children confidently filled the blank with “wugs,” which demonstrates that the children could guess that it is a noun and supplied a standard plural form, suffixing it with an “s.” When someone exhorts you to “can the fat talk and get to the job,”

you will have no difficulty in figuring out that “can” here is not a noun, as in a “coke can,” but it is a verb form of “can” which means to “put something in a can” or “shut something up,” and that “fat” does not refer to lipid levels, but to the unreliable, unsubstantiated manner of your speech. In this context, “can” is a verb, and “fat” an adjective. Thus we observe the truism that we can often surmise the lexical class of an unfamiliar word, or an unusual usage of a familiar word, simply by the company it keeps, its context.

Ritter and Kohonen trained a self-organizing map on short, three-word sentences consisting of nouns, verbs, and adverbs. The words are drawn from a short wordlist and the sentences are constructed by combining these words in several meaningful ways. The tricky part of the training process is figuring out the right representation of a word. Unlike a frequency, which is naturally represented by a single number, the ideal manner of representation of a word is not obvious. Further, a word must not be presented in isolation, but always in its context.

The study used a set of 30 words consisting of nouns (Jim, Mary, dog, horse, etc.), adverbs (seldom, fast, often, etc.) and verbs (eats, hates, sells, etc.). These words are used to form sentences like: “Mary likes meat,” “Jim speaks well” and so on. Each word is represented by a binary seven-dimensional vector. For example, the vector  $x_1$ , which corresponds to the first word Mary, could be

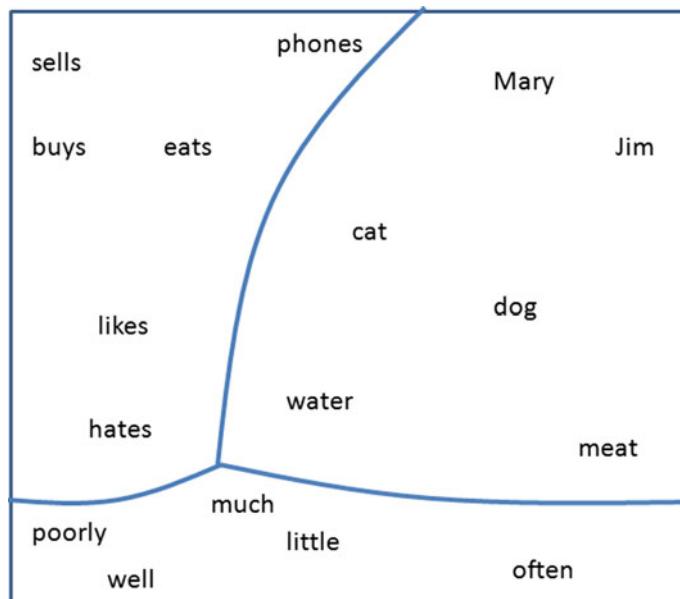
$$X_1 = [1 \ 0 \ 0 \ 1 \ 0 \ 1 \ 0].$$

Each word is accompanied by a context consisting of the preceding word and the succeeding word. Thus, the complete vector,  $X_{1c}$ , which includes the context, for “Mary,” could be something like:

$$X_{1c} = [X_{p1}, X_1, X_{s1}]$$

where  $X_{p1}$  represents the preceding word, and  $X_{s1}$  represents the succeeding word to Mary. Actually, instead of taking individual instances of preceding and succeeding words, for each words, the authors of the study chose to use average of all preceding words (to “Mary”) in place of  $X_{p1}$ , and the average of all succeeding words, in place of  $X_{s1}$ . Ten thousand sentences are constructed from the original word list, and vector representations for each word (similar to  $X_{1c}$  for “Mary”) are computed. The vectors are used to train a self-organizing map of size  $10 \times 15$ . After about 2000 presentations of the training data, it was found that the words are mapped onto the array of neurons in a manner that reflects their semantic and grammatical relationships.

A gross form of ordering that can easily be discerned is that the words are clustered in terms of nouns, adverbs, and verbs (Fig. 6.15). On closer inspection, one may notice that proper nouns (Jim, Mary, and Bob) are closer to each other within the noun region in the map. Thus, proper nouns and common nouns are located at the extreme ends of the noun region. In the adverb part of the map, it is intriguing that opposite adverbs (fast-slowly, little-much) are close to each other. This counterintuitive proximity of adverbs with opposite meaning may be accounted for if we remember that adverbs with opposite meaning can in general snuggle in identical contexts, perfectly replacing each other without affecting the syntax.



**Fig. 6.15** Semantic map of words trained by short sentences (Redrawn based on Ritter and Kohonen 1989)

This and other models of semantic maps constructed using self-organizing principles suggest that the organization of word classes in the cortex is probably governed by similar principles. But modeling results may not be taken as “proof” that self-organizing map mechanisms of the kind described in this chapter actually govern semantic map formation in the brain. Furthermore, the map model just described is too simplistic to be taken as a model of word maps in the cortex, but it provides enough insight and direction to development of more realistic models.

## Discussion

After visiting a small, but important and representative, set of topographic maps in the brain, and seeing how they can be understood, not by invoking a different story in each case, but uniformly by application of a minimal set of mechanisms, we are now ready to ask a basic question about these maps? What are they for? Why does brain need topographic maps? What is their role in brain’s information processing machinery?

The whole excitement about topographic maps perhaps began when Penfield, informed by his pioneering mapping experiments, introduced the idea of homunculus. It seemed to present a picture of the brain, not as an inchoate tangle of myriad components, but a unified, unitary entity processing the influx of sensory information.

It seemed to offer a way of reinstating the person in the brain. It seemed to give back the brain its selfhood. Pictures of homunculus, those odd caricatures of the human form, were widely publicized, and have gained a certain inexplicable explanatory power. The moving, colored image on the retina is found to be mapped, point-to-point on a part of the cortex. It seemed to explain how we see the world. Holes in this map explained holes in our visual perception. The oddly shaped somatosensory homunculus seemed to explain how we receive the touches of the world. It seemed to explain fluctuations in our body image, both minor, like transiently lost sensation of a leg in cramps, and dramatic, like the phantom limb. The idea of a motor homunculus is even more convincing. It seemed to reestablish the idea of a hidden controller, a neural puppeteer that commands our movements.

But at this point, we must hasten to point out that the explanatory power of topographic maps, particularly in their extreme forms as homunculi, is slightly misleading, and a little deeper investigation would lead us to a dead end. Why would the visual system take enormous trouble to reconstruct and maintain a faithful, camera-like image of the external world in the brain, when it was already readily available in the retina? If creation of a world-image in the brain, in the primary visual cortex is the end of the road, what is the function of higher visual areas that are spread out further on in the occipital, parietal and inferior temporal areas? Likewise, if creation of an image of the entire body surface in the somatosensory cortex is the sole strategy of somatic processing, why are there profound fractures in that map? In the somatic map, why do we find the fingers next to the head, genitals next to the toes, and teeth and lips on opposite sides of the lower jaw? Furthermore, there are not one but four different somatic maps on each side of the brain even in the primary somatosensory cortex. Each of these four maps responds to different properties of touch information—light touch, vibration and so on—coming from the skin. Presence of four homunculi seems to defeat the whole purpose of homunculus, contradicting the original need for a single, solitary homunculus. It is even more facile to deconstruct the motor homunculus. Unlike point-by-point maps of somatosensory maps and visual maps of primary visual cortex, the motor maps do not enjoy a strict topographic structure. In detailed maps of hand area in the motor cortex of macaque monkeys, neurons that control digits are in the center, surrounded by neurons controlling wrist. Yonder there are neurons controlling elbow and shoulder. All the neurons that control a given part of the hand are not in one place. There are several patches of neurons that control the shoulder, or the elbow of the same hand. Thus, though at a gross level, there is some sort of a “homunculus,” the microstructure tells a very different story. Moreover, there is not one but several motor homunculi in the many motor cortical areas: beyond the primary motor area, in the higher motor areas like the premotor area and supplementary motor area.

Thus the idea of a homunculus, particularly in its simplistic form, as a “little man” in the brain, is unnecessary and unsupported by facts. But topographic maps, particularly in a coarse sense do exist, though it must be emphasized that what is present is not a separate unitary map for each sensory modality but a complex system of interconnected maps. Therefore, the answer to the question of “why maps” becomes quite tricky. We proceed by answering a related, but more tractable question of “how maps?” and seek to present it as an answer to the former question.

Two mechanisms that we have already presented as forces that shape map formation are Hebbian learning and competition. Hebbian learning achieves amplification of responses, since neurons that respond to a stimulus, initially weakly, strengthen their connections and learn to respond more strongly. Local competition gives us a certain convenient division of labor by which neurons evolve tuned responses to stimuli. Different neurons specialize in responding to different specific stimuli. Both of these mechanisms are obviously desirable characteristics for a nervous system to have. There is a third, and most important, characteristic of map: nearby neurons respond to similar inputs. It is this property that makes the maps topographic. We have discussed possible lateral interactions among neurons, either directly via synapses, or by release of diffusing signals like nitric oxide. There is another possible, key purpose, if you may call it, of topographic maps. Topography may be a result of wire length minimization. It has been shown mathematically that topographic maps could result by evolving response pattern in a sheet of neurons in such a way that wire length is minimized. We have seen in Chap. 2, that wire length minimization is an important evolutionary constraint on the brain. Therefore, topographic maps then seem to be a result of brain's effort to optimize its representations of the world.

## References

- Fishman, R. S. (1997). Gordon Holmes, the cortical retina, and the wounds of war. The seventh Charles B. Snyder lecture. *Documenta Ophthalmologica*, 93, 9–28.
- Fox, J. (1984). The brain's dynamic way of keeping in touch. *Science*, 225, 820–821.
- Gally, J. A., Read Montague, P., Reeke, G. N., Jr., & Edelman, G. M. (1990). Diffusible signal in the development and function of the nervous system. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 3547–3551.
- Jenkins, W. M., Merzenich, M. M., Ochs, M. T., Allard, T., & Guic-Robles, E. (1990). Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *Journal of Neurophysiology*, 63(1).
- Kaas, J. H., Merzenich, M. M., & Killackey, H. P. (1983). The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals. *Annual Review of Neuroscience*, 6, 325–356.
- Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, 6, 895–905.
- Kohonen, T. (1997). *Self-organizing maps*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.
- Martinetz, T., Ritter, H., & Schulten, K. (1988). Kohonen's self-organizing map for modeling the formation of the auditory cortex of a bat. *SGAICO Proc. Connectionism in perspective*, Zürich, 403–412.
- Merzenich, M. M., Nelson, R. J., Stryker, M. P., Cynader, M. S., Schoppmann, A., & Zook, J. M. (1984). Somatosensory cortical map changes following digit amputation in adult monkeys. *Journal of Comparative Neurology*, 224(4), 591–605.
- Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind* (328 p). New York: Quill William Morrow.
- Ritter, H., & Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics*, 61, 241–254.
- Ritter, H., Martinetz, T., & Schulten, K. (1992). *Neural computation and self-organizing maps: An introduction* (revised English edition). New York: Addison-Wesley.
- Suga, N. (1990). Bisonar and neural computation in bats. *Scientific American*, 262, 60–68.

# Chapter 7

## Pathways of Light



*Of all the senses, sight must be the most delightful.*

—Helen Keller.

This chapter is about the problem of vision, which has, interestingly, two subproblems. One of these subproblems is an easy one, the other hard. The easy problem of vision concerns itself with what happens to light when it enters the brain through the portals called eyes. What is its path? What are the major stopovers? What exactly happens at each of these stopovers? The problem is not easy because it is known in all its immense detail. In fact, the details of the visual system are not completely unraveled, despite the intense and sometimes disproportionate attention paid to vision by the neuroscience community. It is easy because the problem is mainly one of getting all the relevant details by expending adequate resources, human and otherwise, and a lot of time. It is easy in the sense that there is a method to go about it. The other problem of vision is not so easy because there is no well-defined method that allows you to make predictable progress in that area. The hard problem of vision deals with the more interesting, popular question: how do we *see*? What are the exact neural events that conspire to enable us to have the moment-to-moment revelation of a moving, multicolored vision of the universe? It is not that neuroscience failed to make any progress in this matter. It is just that this second problem resides on the borders of science and philosophy, leading us on into deeper questions regarding the nature of consciousness and so on. The standard evidence-based methods of science seem to flounder and buckle in tacking the second problem.

This chapter is about the easy problem of vision.

The earliest thinkers of vision were divided into two camps. One camp held that when we see, the eyes actually emit something that illuminates the world around, a theory known as the *emission theory*. Ancient Greek mathematician Euclid was a proponent of this theory, which is rather unexpected because Euclid actually studied light systematically. He knew that light travels in straight lines. He understood the laws of reflection. In spite of these insights, he thought that vision is a result of the

light emitted by the eyes. Greek astronomer Ptolemy also was an adherent of this theory. The rival theory known as the *intromission theory* believed that something came *from* the object *to* the eye, enabling perception. This view was supported by philosopher Aristotle and physician Galen, both of ancient Greece. In what may be described as one of the earliest classics on Optics, the Arab philosopher Al-Haytham presented rational arguments supporting intromission theory. Al-Haytham argued that since prolonged and direct exposure to sunlight can actually harm the eye, it cannot be true that light is emitted by the eye. He also reasoned out that it is highly improbable that the light emitted by this tiny organ is adequate to fill the heavens and give us a vision of the sun, the planets, and the stars.

We now know that eyes do not emit light, and actually go to great lengths to retain and make careful and niggling use of the photons that enter them from the world without, as we will see later. The essence of the eye's function is to convert the light entering it into electrical signals and convey those signals to the brain via nerve fibers. Response to light is not a special virtue of the eye. In fact, it is ubiquitous in the biological world. Plants use the energy in the sunlight to transform carbon dioxide and water into starch, by the familiar process of photosynthesis. Response to light is what drives the sun-tracking movements of a sunflower.

## Shaping the Eye

Evolution has certainly come a long way in taking this light sensitivity as a seed mechanism and shaping a complex and exquisite organ called the eye. From the state of a featureless tissue responding to photons, to an organ pair extolled as "the windows of the soul," it is an immense progression. Even Charles Darwin, the great proponent of the theory of evolution, hesitated for a moment and wondered how something as primitive as natural selection could have evolved (designed?) something as delicate and complex as the eye. But then he comes to terms with it in these terms in his *Origin of the Species* (p. 172):

...if numerous gradations from a simple and imperfect eye to one complex and perfect can be shown to exist, each grade being useful to its possessor, as is certainly the case; if further, the eye ever varies and the variations be inherited, as is likewise certainly the case and if such variations should be useful to any animal under changing conditions of life, then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, should not be considered as subversive of the theory.

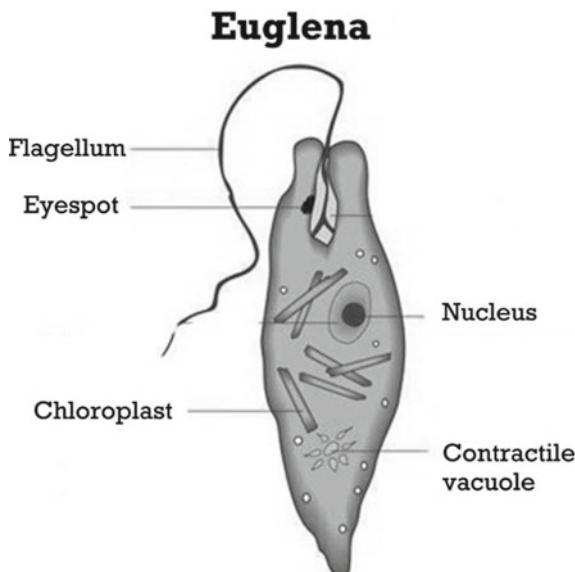
Since the human eye is such an exquisite organ, inspiring so many features of a modern camera, and more, it is tempting to imagine that mere natural selection may not be adequate to shape such a wonder; some sort of "design" may be involved. Although he begins with a certain skepticism about this matter, Darwin strongly expresses the conviction that the evolution of the eye from a simple biological photoreceptor to what it has become today, can be reduced to a large number of intermediate stages which could be feasibly driven by the force of natural selection.

A delineation of exactly those putative “numerous gradations” was attempted by Nilsson and Pelger in their 1994 paper on possible stages in the evolution of the human eye. In this oft-quoted paper, the authors describe and define—not one, not two—but 1829 hypothetical evolutionary steps of the eye.

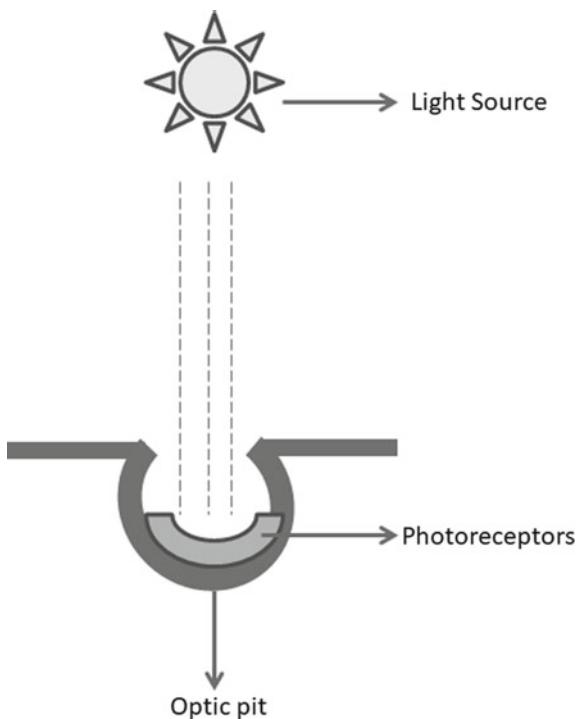
Some of the broad intermediate stages in the Nelson and Pelger account, irrespective of their evolutionary relevance, give an insight into the making of the eye. The most primitive eyes could be nothing more than a patch of light-sensitive cells pasted on the body surface of a creature. There are indeed creatures with such light-sensitive patches called *eyespots*. For examples, some flatworms have eyespots that are sensitive to both light and chemicals. Other examples include caterpillars, starfish, and segmented worms. Some organisms whose bodies are covered with eyespots actually have mostly translucent bodies. In such organisms, the eyespots are not sufficient to detect the direction of the light source. They can only help the host organism to determine the overall illumination levels in the ambience. However, in some protozoa like the flagellates, the eyespot and the flagellum, a whip-like structure whose lashing movements propel the organism through a fluid medium, act as one coordinated unit, driving it toward light (Fig. 7.1).

In the next broad evolutionary stage of the eye, the light-sensitive patch dimples inward forming a cup. Light-sensitive cells, present at the bottom of the cup, are now in a position to detect light coming only in one direction, thereby producing a direction-sensitive response to light. Such cup-like features, known as *optic pits*, are found in certain types of mollusks (Fig. 7.2). The subsequent stage in the story of the growing eye is a further deepening of the depression seen in the last stage, in such a manner that the epidermal edges surrounding the mouth of the cup close

**Fig. 7.1** Eyespot shown on Euglena, a flagellate



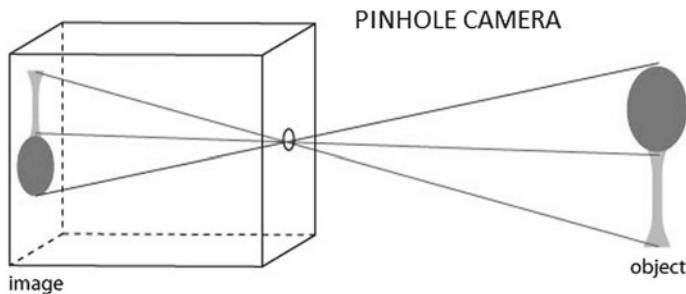
**Fig. 7.2** A schematic of an optic pit



over forming a vesicle. Such camera-type eye with a narrow aperture is found in the cephalopod *Nautilus*.

A stage in the evolution of the eye where we have essentially a closed enclosure with a tiny aperture marks a very important stage. It must be noted that the foremost function of the eye is not just to detect light but to *form an image*. A flat sheet of photoreceptors exposed to the world or a cup-like structure with photoreceptors at the bottom, can detect light intensity, and perhaps even its direction; but it cannot form an image. An optical device must satisfy a special criterion to be able to form an image.

A flat screen is not going to form the image of the object in front of it, no matter how strongly the object is illuminated. Presence of an image implies a correspondence between the external scene and the image. For an image to form, light from a certain point on the external scene must impinge precisely at a corresponding point on the surface that bears the image. On a flat screen directly exposed to the scene, light from every point on the scene impinges on every point on the screen, though at different angles. A simple way to achieve a correspondence between the external scene and the light on the screen, is to constrain the light to pass through a *pinhole*. Making a pinhole camera, with a pinhole on one side of a shoebox and a translucent screen on the opposite side acting as a screen, can be an engaging science project for children (Fig. 7.3).



**Fig. 7.3** A schematic of a pinhole camera

The fact that an enclosure with a pinhole can be used to form an image has been known since the antiquity. In his classic treatise on Optics (*kitab al-manazir*), ancient Arabic astronomer–mathematician Al-Haytham (965–2040 AD) wrote about image formation using the pinhole effect and likened the human eye to a pinhole. Italian scholar and playwright Giambattista della Porta (1535–1615) wrote of the pinhole and noted that the image formed is upside down. On occasion, images formed by the pinhole effect can even be observed in nature. At the time of a solar eclipse, gaps among the leaves of trees can serve as pinholes, painting beautifully arrayed images of the crescent sun on the ground.

In the next stage in the evolution of the eye, as outlined by Nilson and Pelger, the vesicle or sac which is the precursor to the eye, is filled with a jelly-like fluid. Presence of such a fluid in the pit can have multiple benefits. It can help the sac maintain its shape; it can protect the light-sensitive cells at the bottom by keeping germs and dirt away. Retention of the fluid in the sac requires the presence of a membrane closing the mouth or aperture of the sac, a feature that might lead to a possible jump to the next stage in the eye's evolution.

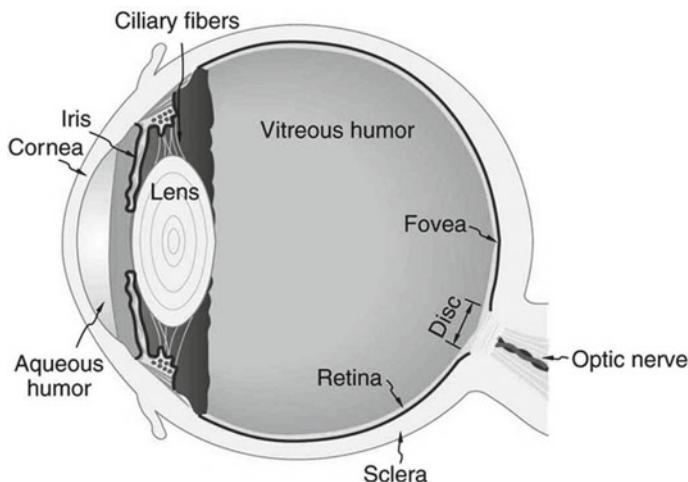
The membrane that closes the aperture must necessarily be transparent. Now, consider a repetition of what happened to the eyespot on the flat surface now happening to the recently formed membrane. The membrane too dimples inward, and forms another sac filled with a transparent fluid. Due to the presence of the second sac, light entering the eye, or its version that has evolved so far, ends up crossing two borders—one between the air and the outer sac, and the second between the outer sac and inner one. The recently introduced outer sac can act as a lens, refracting light and focusing it on the sheet of light-sensitive cells. By suitable adaptations, the outer sac probably transformed itself into a biconvex lens, the likes of which are found even in a segmented marine worm like the polychaeta. The lens can be rendered more effective by increasing the refractive index of the fluid it contains. The greater the refractive index, the greater the focusing of the image.

## Capturing the Image

At the end of the brief evolutionary outline just narrated, we have an eye that is essentially a ball with an aperture or a pinhole through which light can enter. Inside the ball, on the wall opposite to the aperture, there is a layer of photosensitive cells on which the image of the external world forms. Near the aperture there is a lens that focuses the light entering, thereby creating a sharp image on the photosensitive layer. What we have now is a cartoon picture of the human eye, which has actually a much richer architecture and a lot greater sophistication.

The human eye is ensheathed in a fibrous coat called the fibrous tunic. The part of this coat that overlays the lens—the *cornea*—bulges out a bit and is transparent, allowing entry of light into the eye (Fig. 7.4). The rest of the coat is white, constituting the *sclera* or the white of the eye. By virtue of its convexity, the cornea is also capable of bending and focusing light. In fact, 70% of the task of focusing light in the eye is done by the cornea. The space between the cornea and the lens is filled by a thick, gelatinous, and colorless fluid called the aqueous humor while the one behind the lens, filling the ball of the eye, is called the vitreous humor. There is definitely nothing funny about these fluids—the terminology is a hangover from ancient Greeks who referred to body fluids as “humors.”

Since the function of the eye is to receive and convert light, it needs to operate under a very important stringent constraint: the parts of the eye that fall on the path of the light must necessarily be transparent. The constraint explains the transparent nature of the two humors in the eye. The lens too is transparent for the same reason. But the transparency brings with an additional difficulty of supplying nutrients to the lens. Nutrients are supplied to body’s tissues through the blood in the vascular



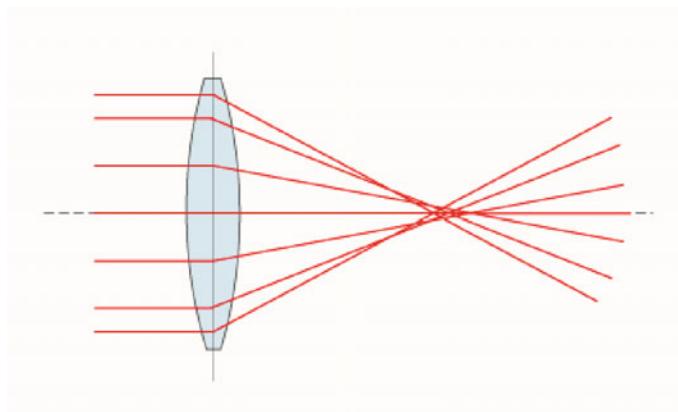
**Fig. 7.4** Anatomy of the human eye

network. But the colored infrastructure of the vascular network is unsuitable for the tissue of the lens. Therefore, the colorless aqueous humor serves as an intermediary between the vascular network and lens in carrying nutrients to the lens. The aqueous humor has another important purpose—the pressure it exerts on the cornea helps to maintain the bulge of the cornea.

Overlying the lens, partly covering it, there exists a ring of muscles called the *iris*. At the center of the iris there is a hole called the *pupil* that controls the amount of light entering the eye. Two sets of muscles control the size of the pupil. Contraction of one of these muscles constricts the pupil, while the contraction of the other dilates it. Muscles of the iris, constrictor pupillae, dilator pupillae, controlled by parasympathetic and sympathetic system, respectively, the two main branches of our involuntary or *autonomous* nervous system. Pupils involuntarily dilate when we express interest in an object or a person. This simple physiological fact is the basis of the social impression that wide-open pupils render their owner more attractive. Italian women of the middle ages used a plant extract called *belladonna* (Italian for a “pretty woman”) as a part of their self-beautification ritual. Belladonna contained a chemical called atropine that dilates pupils. This unconscious tendency to associate dilated pupils with attractiveness has also been studied in controlled conditions. In 1965, psychologist Eckart Hess presented to subjects two sets of images of women: one of women with average-sized pupils and the other with dilated pupils. The viewers, all men, found the images with dilated pupils more attractive.

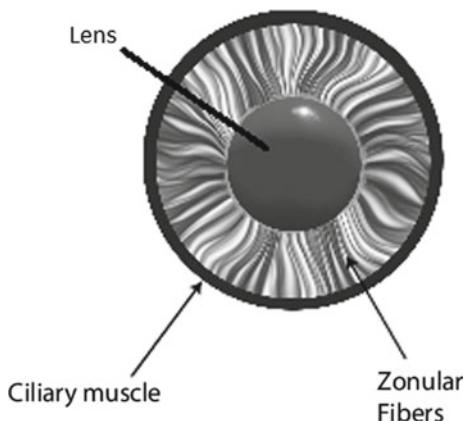
The lens has other sophisticated features like, for example, its heterogeneous internal structure. Ordinary lenses, of the kind found in a high school physics lab, are made of homogeneous material, and therefore have a uniform distribution of refractive index. Moreover, such lenses have typically spherical surfaces, since they are cheaper to produce. Spherical lenses suffer from a drawback called spherical aberration: since light entering the lens near the periphery is bent more than light entering close to the center, a parallel beam does not converge perfectly at a focal point (Fig. 7.5). This can be fixed using an inhomogeneous refractive index—high in the center, falling toward the periphery. With such a distribution of the refractive index, light entering near the center is also bent sufficiently so as to produce a precise focus. The lens in the eye consists of layers of transparent cells containing a protein called *crystallin*. The concentration of crystallin in a given part of the lens controls the local refractive index. Thus, the distribution of the refractive index in the lens is controlled by adjusting the distribution of crystallin.

The thickness of the lens is also adjustable by the joint and complementary action of two anatomical features—the ciliary muscles and suspensory ligaments called zonules. Zonules are ligaments surrounding the lens, pulling them radially outward. Normally, zonules are in tension, and therefore pull the lens outward flattening it. A flat lens has a longer focal length and is ideal for viewing objects at a distance. The ciliary muscles surrounding the lens are relaxed when the zonules are taut. Contrarily, when the ciliary muscles contract, the zonules relax, thereby allowing the lens to bulge. A swollen lens has a shorter focal length and is ideal for viewing nearby objects (Fig. 7.6). This ability of the lens to dynamically adjust its focal length according to the target distance is called *accommodation*.



**Fig. 7.5** Spherical aberration caused by the spherical surface of a lens

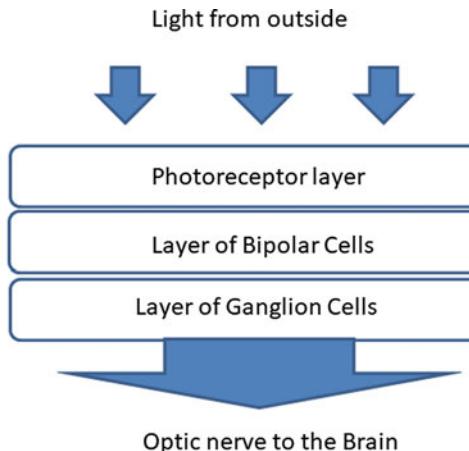
**Fig. 7.6** Muscles that control a lens



Light, thus squeezed by the pupil, bent to shape by the cornea–lens system, is focused on the screen called the retina located on the inner wall of the eye opposite to the pupil. The sheet of the photoreceptors that converts the received light to electricity is located inside the retina, which is a lot more than just photoreceptors. The retina actually consists of two layers of neurons in addition to the layer of photoreceptors. The electrical signals produced by the photoreceptor layer are first processed by the two neural layers embedded inside the retina before they are transmitted to the brain via the optic nerve.

Let us now visit the retina where the beginnings of neural image processing take place. Since we already mentioned that there are three layers in the retina—the layer of photoreceptors and two neural layers—it is absolutely imperative that light first hits the photoreceptor layer, the output of which is processed by the neural layers in some sequence. To enable such a sequence in processing steps, it is logical to

**Fig. 7.7** The verted configuration of retinal layers



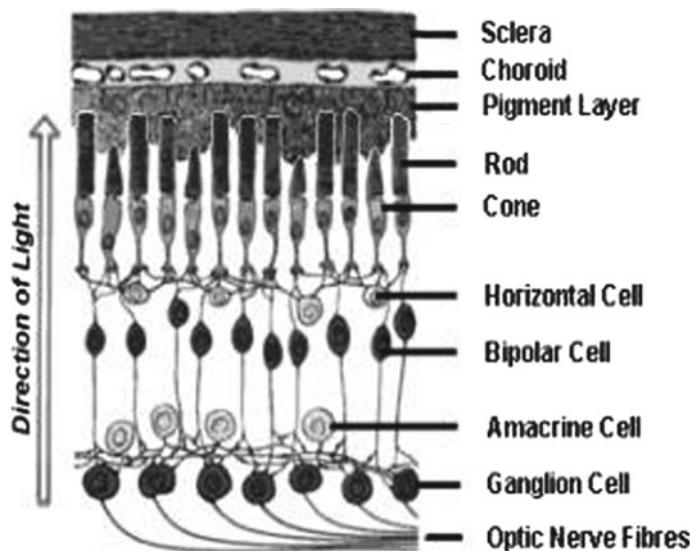
#### LOGICALLY EXPECTED ORDER OF THE RETINAL LAYERS

assume that the three layers are placed in the following order facing the impinging light (Fig. 7.7).

Paradoxically, the actual arrangement of the three layers in the retina of the human eye is exactly opposite to what seems eminently logical. This actual arrangement is said to be verted, and therefore, the seemingly more logical arrangement (Fig. 7.7) is said to be just verted. In the verted configuration of Fig. 7.7, light falling on the retina first hits the photoreceptors directly. In the verted configuration, light has to cross two layers of neural tissue before it reaches the layer of photoreceptors. There is a controversy surrounding this counter-intuitive arrangement of retinal layers. Noted evolutionary biologist Richard Dawkins expresses a certain ambivalence in coming to terms with this situation.

With one exception, all the eyes I have so far illustrated have had their photocells in front of the nerves connecting them to the brain. This is the obvious way to do it, but it is not universal. The flatworm keeps its photocells apparently on the wrong side of their connecting nerves. So does our own vertebrate eye. The photocells point backwards, away from the light. This is not as silly as it sounds. Since they are very tiny and transparent, it doesn't much matter which way they point: most photons will go straight through and then run the gauntlet of pigment-laden baffles waiting to catch them.

Tolerating the above paradoxical situation with a conciliatory attitude of “it doesn’t matter which way they point” is not the same as searching for possible advantages in this seemingly “verted” arrangement of retinal layers. There have been efforts to account for this paradox in terms of the peculiar metabolic needs of the photoreceptors. The metabolic rates of the retina are highest compared to any other tissue in the body. Although at the whole organ level, the brain has the highest metabolic rates, at the tissue level, the metabolic rate of the retina is 300% greater than that of the cortex. A big part of this metabolic demand in the retina comes from the photoreceptors. In the retina, microvessels run along a layer of tissue called the choroid layer, a layer



**Fig. 7.8** Layers of the retina

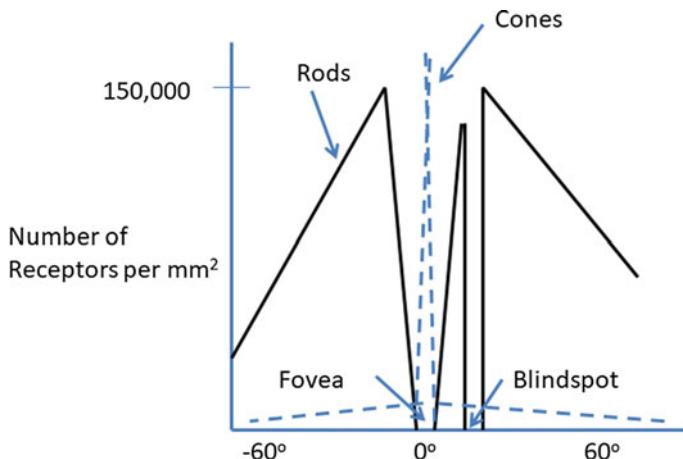
that is in close proximity to the photoreceptors. Thus, in the so-called verted configuration, the photoreceptors, by virtue of their close proximity to the blood vessels, get to meet their extraordinary energy demands.

Perhaps another advantage can be linked to this proximity of the photoreceptors to the vascular bed. As the photoreceptors perform the operation of converting light into electricity, they constantly shed disk-like parts of themselves (called “segments”) while new ones are constantly added. The debris created by all the shedding needs to be cleared, which is where the circulatory system comes in.

Yet another purpose can be seen behind the microanatomical fact that the photoreceptors are deep inside the retina, and not on the surface. The inner walls of the eye, other than the side where the lens is present, are lined by pigmented cells known as Retinal Pigment Epithelial (RPE) cells. They contain a pigment known as melanin, which incidentally gives the skin its darker shades. In darker skins, melanin absorbs light. In the verted configuration, the photoreceptors are proximal also to the RPE cells. Light that is not absorbed by the photoreceptors will be scattered around. If the scattered light is not mopped up, we will be seeing the world in an intolerable whitish haze. Mopping up the scattered light seems to be the job of the melanin-containing RPE cells (Fig. 7.8).

Considering the above advantages, the verted arrangement is not just better; it is probably inevitable.

Let us now consider the photoreceptors, whose presence defines the very reason to be of the eye. Broadly speaking there are two kinds of photoreceptors—rods and cones—in view of their shapes. There are about 120 million rods and about 7 million cones in each eye. Rods and cones are not randomly distributed over the



**Fig. 7.9** A simplified depiction of the distribution of rods and cones in the retina

retina, like salt and pepper, but have characteristic distributions with close relevance to their respective functions. In the center of the retina, there is a region, about half a millimeter in diameter, called fovea. Although no rods are present, the fovea has a high concentration of cones; elsewhere in the retina, rods are more prevalent. Thus, cones and rods have roughly complementary distributions over the retina (Fig. 7.9). Cones are present at high concentrations in the fovea and not much outside the fovea; rods are present in high concentrations everywhere except the fovea. The functions of these two sets of photoreceptors also, reflecting their spatial distribution over the retina, are complementary.

Rods are excellent at detecting very weak intensities of light; at stronger intensities, their response hits a ceiling and is no more informative. Therefore, rods swing into action under the conditions of low ambient light as, for example, in a dim lit room or on a starlit field at night. Rods are found to be so sensitive that they are capable of responding to single photons. Cones, on the contrary, respond to stronger intensities and do not saturate. Therefore, they are suitable for daylight vision. Another advantage of cones is that they respond to colored light. There are three kinds of cones sensitive roughly to red, blue, and green wavelengths. Since cones do not respond under conditions of low ambient light, color perception occurs only when the ambience is sufficiently bright. Thus, we see a splendid division of labor between rods and cones—one for the night vision and the other for bright vision. In fact, this philosophy of minute and microscopic division of labor seems to be one of the fundamental design philosophies of the brain. One can identify innumerable cases in neurobiology where single neurons do something extremely specific. We have visited this idea of specialized functions of neurons in Chap. 6. Likewise, as we progressively describe the hierarchy of the visual system, we will recognize the power of this minute division of labor, with neurons in various areas of the visual system specializing in very specific aspects of visual processing.

The outputs of the photoreceptor layer are processed by two layers of neurons before visual information exits the retina and enters the brain. The first of these neural layers, the layer of bipolar cells, receives the outputs of the photoreceptors. The second neural layer known as the ganglion cell layer, in turn, receives the outputs of the bipolar cell layer. Within the layer of bipolar cells, there is a class of neurons called the horizontal cells that enable interaction among the bipolar cells. Similarly, there are neurons called amacrine cells that allow interactions among ganglion cells. The axons of the ganglion cells collect at a point in the retina where they exit as a bundle known as the optic nerve. Due to the peculiar features of the verted configuration, this collection of axons of ganglion cells at a point in the retina creates a local difficulty. Since the ganglion cells are close to the surface of the retina, their axons creep along the surface of the retina and meet at a point where they punch through and exit the retina. Therefore, at the point where the axons exit the retina, there is no place for photoreceptors. For this reason, light falling on this spot of the retina cannot be sensed, thereby producing a *blind spot* in the eye (Fig. 7.9). Note that the blind spot could have been avoided in an verted configuration.

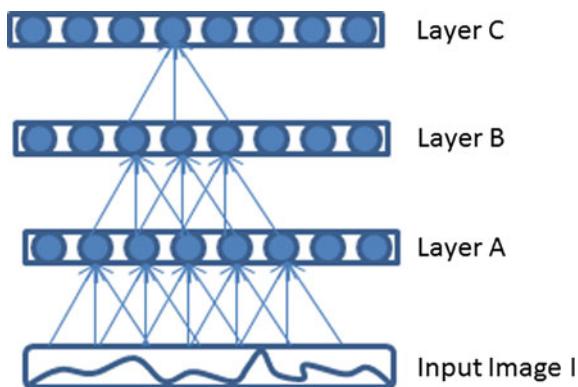
We are now ready to familiarize ourselves with an idea that is crucial to understand the organization of the visual system—*the receptive field*. A tiny spot of light shone at a point on the retina obviously produces responses, if at all, only in neurons close to the spot. Thus, the receptive field of a neuron consists of the part of the visual field in which a stimulus, say, a beam of light, can produce a response in the said neuron. Let us now discuss the idea of a receptive field first in more abstract terms, using a general multilayer neural network as a prop. We will then apply the idea to the specific organization of the visual system.

Consider a general neural network with multiple layers of neurons—A, B, C, and so on (Fig. 7.10). The input layer I, which is analogous to the layer of photoreceptors, carries information about an image which is presented as input to the first neural layer A. In this simple network architecture, neurons in each layer project successively to neurons in the subsequent layer, just as in a multilayer perceptron of Chap. 4. However, the present network differs from a multilayer perceptron in a crucial way. Each neuron in a given layer receives inputs, not from all the neurons of the previous layer, as in the case of a multilayer perceptron, but only from a small “window” of neurons in the previous layer. For example, in the schematic of Fig. 7.10, each neuron in layer A receives inputs from three photoreceptors in I. Likewise, each neuron in B receives inputs from three adjacent neurons in layer A and so on.

Now consider a spot of light that activates three adjacent photoreceptors, I3, I4, and I5 in I (Fig. 7.10). I3, I4, and I5 are precisely the photoreceptors that project to neuron A4 in layer A. A4 does not receive inputs from any other photoreceptors. Thus, I3, I4, and I5 constitute the receptive field of neuron A4. The spot of light that activates I3, I4, and I5 is likely to produce a response in neuron A4. Similarly, I2, I3, and I4 constitute the receptive field of neuron A3; I4, I5, and I6 constitute the receptive field of neuron A5. Thus, as we move from neuron to neuron in layer A, the corresponding receptive field also shifts in the photoreceptor layer.

Now consider the receptive fields of neurons in layer B. Consider neuron B4 which receives inputs from neurons A3, A4, and A5 in layer A. The receptive fields of these

**Fig. 7.10** A multilayer neural structure that reveals the notion of a receptive field

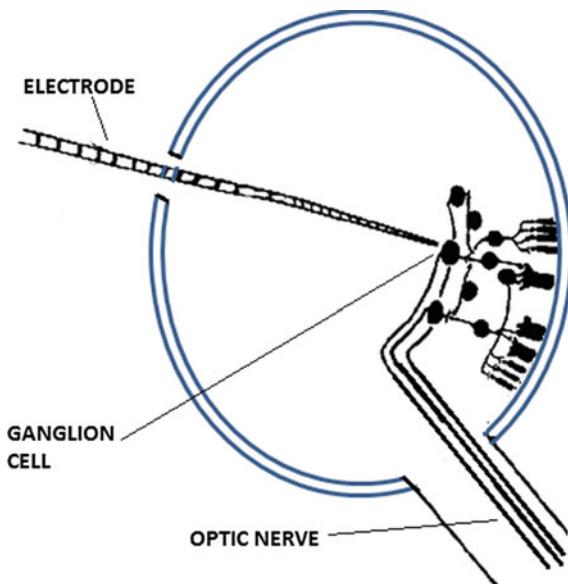


three neurons in turn are: {I<sub>2</sub>, I<sub>3</sub>, I<sub>4</sub>}, {I<sub>3</sub>, I<sub>4</sub>, I<sub>5</sub>}, and {I<sub>4</sub>, I<sub>5</sub>, I<sub>6</sub>}, respectively. A light spot activating any of the five photoreceptors (I<sub>3</sub> to I<sub>7</sub>) is likely to produce an activation in B<sub>4</sub>. To put it more precisely, there are anatomical pathways that connect neuron B<sub>4</sub> to the photoreceptors I<sub>2</sub> to I<sub>6</sub>; B<sub>4</sub> is not connected to any other photoreceptors. Thus, photoreceptors I<sub>2</sub> to I<sub>6</sub> constitute the receptive field of B<sub>4</sub>. It is noteworthy that while the size (# of photoreceptors) of the receptive fields of neurons in layer A is only three, the size of those in layer B is five. Thus, with the kind of pattern of projection seen in the network of Fig. 7.10, neurons in higher layers have larger receptive fields. The higher up a neuron, the more it gets to see.

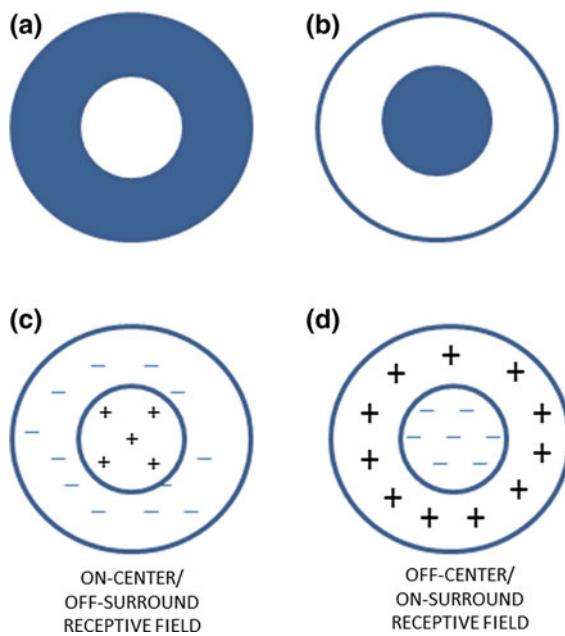
Now let us return to the story of real receptive fields of the retinal neurons. The earliest work on the receptive fields of the retinal neurons was done on the ganglion cells by H. K. Hartline at Rockefeller University, USA, Stephen Kuffler at Harvard University, USA, and Horace Barlow at Cambridge University, England. In these pioneering studies, a small spot of light was flashed on the retina while recordings were made from an electrode inserted into a ganglion cell. The schematic of Fig. 7.11 shows how such a recording is made. The spot was moved around until it produces a response in the ganglion cell from which the recording is made. The response is a sign that the spot is inside or is overlapping with the receptive field of the neuron.

Once the whereabouts of the receptive field of a neuron are found, the search is directed toward finding the exact pattern of light stimulus that maximizes the response in the neuron. To this end, the experimenters varied the spot size and found the spot size at which the neurons show the strongest response. On more careful probing, the experimenters found that best response was obtained not by a spot of uniform luminance but one in which there is a contrast between the central region and the surrounding portion of a circular visual stimulus. Since contrast can be of two types—black-against-white or white-against-black—two types of stimuli are typically found to produce strongest responses in ganglion cells (Fig. 7.12a,b). Ganglion cells were also begun to be named after the stimuli that produce best responses in them. Neurons that respond to stimuli with a central bright spot and surrounding darkness are said to have ON-center/OFF-surround receptive fields (Fig. 7.12c). Stimuli

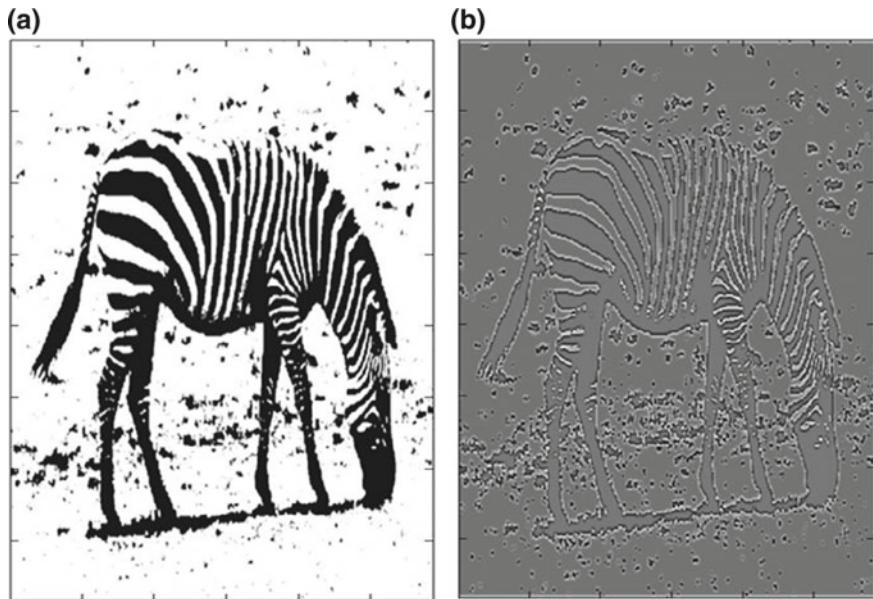
**Fig. 7.11** A schematic of the experimental setup used to record from the retinal ganglion cells



**Fig. 7.12** ON-CENTER/OFF-SURROUND and OFF-CENTER/ON-SURROUND receptive fields



with any other distribution of light and dark produce weaker responses. Contrarily, there are also ganglion cells with OFF-center/ON-surround type of receptive fields (Fig. 7.12d).



**Fig. 7.13** An illustration of the expected response of ganglion cells with balanced CENTER-SURROUND receptive fields. For the zebra picture in (a), the response consists predominantly of the edges (b)

An interesting question about the meaning of ganglion cell responses can be raised at this point. Although the photoreceptors simply respond to intensity in light, though with some selectivity in frequency, why do ganglion cells respond to such a specialized pattern? Note that a ganglion cell responds best when there is a balance of light and dark in its receptive field. Its response to a stimulus that is all bright or all dark is not significant. Thus, ganglion cells respond better when there is a contrast in the image shone in its receptive field.

Let us now consider, hypothetically, the nature of the response of ganglion cell layer to black and white image like that of the zebra in Fig. 7.13. The grayish image in Fig. 7.13a depicts the simulated response of simplified models of center-surround cells. Both white and black regions in the input image are evened out to gray. But the edges between the black and white, where there is obviously a greater contrast, are amplified in the ganglion response image (Fig. 7.13b). Thus, ganglion cells seem to be enhancing contrast in the retinal image and emphasizing edges. But the presence of such mechanisms at an early stage in the visual system is only natural since both contrast enhancement and detection of edges are some of the most primitive operations that computer scientists apply to images from which they seek to extract information about various objects contained in the image and their identity.

At this point, one must hasten to note that contrast is not the only information that is conveyed by the ganglion cells. Or, rather, it is not just the contrast between black and white since the ganglion cells receive, via bipolar cells, information about

moving form in multicolor. Information about color and also motion in the retinal image is conveyed by the ganglion cells to the next visual stopover in the brain.

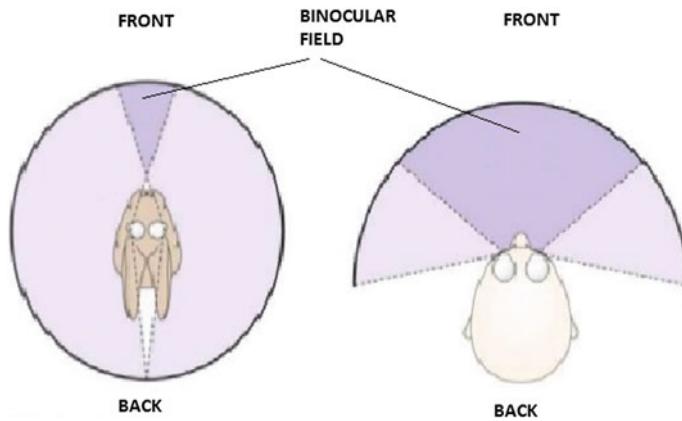
What the eye or the retina conveys to the brain is the output of the ganglion cells. Axons of the ganglion cells form a bundle called the optic nerve which projects to a part of the thalamus known as the Lateral Geniculate Nucleus (LGN). There are two optic nerves, one from either eye. Thalamus (actually, there are two thalami located bilaterally in the two hemispheres) is an important hub through which most sensory information streaming into the brain from the world must pass.

How do the two optic nerves connect to the LGNs of the two thalami? Does the optic nerve emerging from one eye project to the LGN on the same side of the brain or to the opposite side? Actually, the pattern of projection of optic nerves to the two LGNs is quite complex. For each optic nerve, one part of the fibers projects to the LGN on the same side, while the other part projects to the LGN on the opposite side. The pattern of this projection is so intricate and precise, with individual optic nerve fibers projecting to very specific targets in the neurons of LGN. These complex wiring patterns are a marvel of neural development, comparable in their intricacy to the wiring patterns in a modern VLSI chip.

In order to understand the logic of connectivity between the eye and LGN, one must first consider how the two eyes share the labor of processing different parts of the visual field.

Figure 7.14 shows how our eyes are placed with respect to what is ahead of them, the visual field. Although the two eyes see nearly the same scene, what they see is not identical. This is understandable because each eye is at a slightly different vantage point. Therefore, there are bound to be subtle differences in the retinal images captured by the two eyes. In fact, these minute differences in the two retinal images are partly the basis for our depth perception. But there is a more gross difference between what the two eyes see. There are parts of the visual field that can be seen by both the eyes and parts that are exclusively visible by a single eye. For example, what the left eye can see stretches from the leftmost point on the arc with a lighter shade, and extends to the right side, falling slightly short of the rightmost extreme. The left eye cannot see that small stretch of the visual field at the right extreme because the nose blocks the view. Similarly, the visual field visible to the right eye can also be identified in the same figure. Barring the two small sectors at the extremes, there is a large common portion of the visual field that is visible to both the eyes. This part is known as the binocular (= “both eyes”) field. The part on the left extreme visible to the left eye alone is called the left monocular field and the part on the right visible to the right eye alone is the right monocular field. On the whole, the monocular fields are smaller than the binocular field. But that is particularly true in humans and not necessarily in animals.

The prominence of the monocular fields relative to binocular field depends on where the eyes are positioned in the head. There are animals, like us, in whom the eyes are placed in the front of the head. In such cases, as we have just seen, the binocular field is large. But there are others in whom the eyes are to the side of the head. Horses and hares, dogs and deer are some examples. With the eyes located to the side, these animals have large visual fields: they can see what is to their sides, and



**Fig. 7.14** The binocular field in animals (left) and humans (right)

**Table 7.1** The mapping between the hemiretinas and the (left and right) visual fields

| Hemiretina                | What it sees       |
|---------------------------|--------------------|
| Left nasal hemiretina     | Left visual field  |
| Left temporal hemiretina  | Right visual field |
| Right nasal hemiretina    | Right visual field |
| Right temporal hemiretina | Left visual field  |

to a great extent behind them too (Fig. 7.14, left). Thus, they possess large monocular fields on the sides and a relatively small binocular field in the front. Animals with sideways facing eyes can spot predators as they creep up behind them. But they have poorer depth perception which requires the two eyes to be seeing very similar, but not identical, scenes. Thus, depth perception is stronger in heads with eyes in the front.

There is yet another useful way of dividing the visual field. The left visual field and the right visual field denote simply the left and right halves of the visual field, respectively. The left (right) eye can see the left (right) visual field completely and a major part of the right (left) visual field. Therefore, either eye can see both halves of the visual field almost entirely. On this basis, the retinas of the two eyes are also divided into two halves—the hemiretinas. We could have kept the jargon simple and called these halves “left” and “right” hemiretinas. But since neuroscientists prefer to call them “nasal” (on the side of the nose) and “temporal” (on the side of the temples or ears) hemiretinas (half-retinas), we will swallow the pain for the moment and adopt this jargon. Since light enters the eye through the narrow aperture of the pupil, the left (right) part of an eye gets to see the right (left) visual field. This left-right inversion happens in both the eyes. For this reason, the visual field gets split among the four hemiretinas (of the two eyes) in a manner neatly described by Table 7.1.

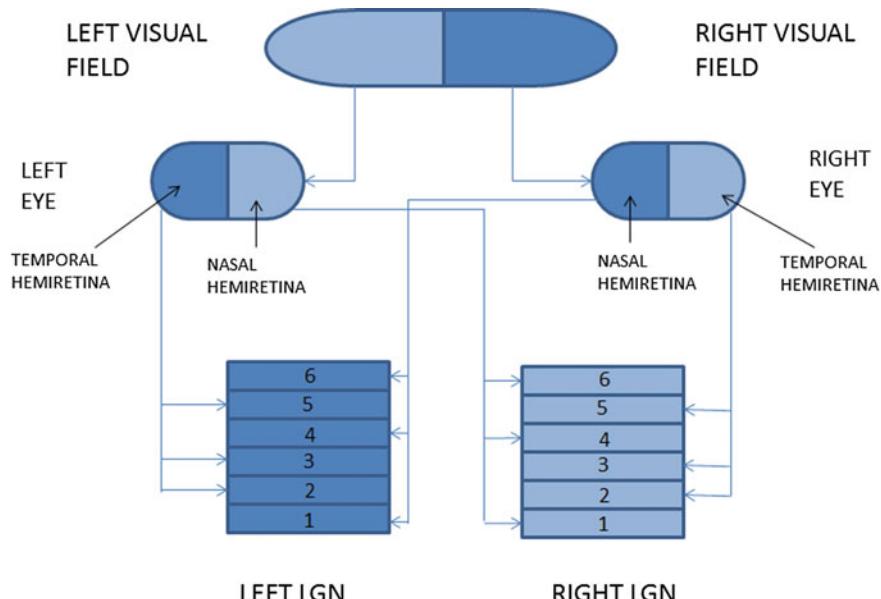
The long drawn description of how the visual field and the two retinas are labeled is essential to understand how information from the eyes is projected to the brain. There is a general strategy that brain follows in mapping the visual space onto the visual cortex. Similar strategies are used by other sensory systems also, as we will consider in subsequent chapters. The left (right) visual field is mapped onto the visual cortex of the right (left) hemisphere, or, the right (left) visual cortex. But since information about the left and right visual fields is present in both the eyes, it requires some really intricate wiring between the eyes and the brain for such a mapping to be possible.

The wiring from the eyes to the brain is configured as follows. Since the temporal side of either eye looks at the opposite visual field, optic nerve fibers from the temporal sides project onward to the same side of the brain. And, since the nasal hemiretina of either eye looks at the visual field on the same side as the eye, fibers from here project to the opposite side of the brain. Fibers that travel to opposite side of the brain crossover at a junction point called *optic chiasm*. (Chiasm is simple “neuroanatomese” for crossing over.) Beyond the optic chiasm, fibers from the two eyes are mixed; but both sets of fibers now carry information from the same visual field.

The first stopover of these fibers in the brain is the thalamus, a major port of entry for most sensory information entering the brain. The thalamus is like the reception desk of a hotel where the guests are routed to their respective rooms. As mentioned before, a part of the thalamus called the lateral geniculate nucleus, in charge of routing visual information, receives the optic fibers from the two eyes. These fibers are connected in intricate ways to various subdivisions in the LGN.

The six layers of LGN on each side of the brain are numbered from 1 to 6. Figure 7.15 shows the connections from the eyes to the LGNs. Since the purpose of rewiring is to segregate information related to the right versus the left visual field, regions that deal with left and right visual field information are colored gray and white, respectively. Optic fibers from the eye on the same side as the LGN are directed to layers 2, 3, and 5. Fibers from the eye on the opposite eye are directed to layers 1, 4, and 6. Thus, all the layers in the LGN receive information about the opposite visual field. It is just that three layers receive that information from the eye on the same side, while three others from the opposite eye.

But then since there only two categories (same side versus opposite side) why should LGNs have six layers each? Two layers should have been sufficient. The truth is there are more categories that need to be segregated. We have seen that there are two kinds of ganglion cells. Those that have ON-center, OFF-surround (ON/OFF) receptive fields respond to bright dots on a dark background. Those that have OFF-center, ON-surround (OFF/ON) receptive fields respond to black dots on a bright background. This constituted the second category. The ON/OFF cells project to two of the layers numbered 3–6; the other two layers receive projections from OFF/ON cells. There is another way of classifying the ganglion cells. Some ganglion cells respond to moving form, while others respond to color and not much to movement. The movement-sensitive ganglion cells project to layers 1 and 2 of the LGN, while the color sensitive ones project to remaining layers (3–6). In summary, though the



**Fig. 7.15** Connections from the hemiretinas and the LGNs in either hemisphere

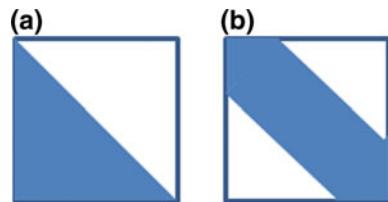
segregation of optic fibers in terms of the eye from they originate is clean, segregation of the other two categories is more mixed in LGN.

In spite of all the intricate mapping of the projections from the eyes, LGN neurons too have receptive fields with simple ON/OFF structure similar to those of ganglion cells. But the situation changes radically when we move on to the next stopover in these pathways of light—the primary visual cortex.

## The Primary Visual Cortex

The primary visual cortex, lovingly called the V1, is the first cortical processing area for visual information coming from the eyes. Pioneering studies conducted by David Hubel and Thorsten Wiesel at Harvard University in late 1950s brought to light some of the fundamental aspects of the function of primary visual cortex. In 1958, Hubel and Wiesel were performing recordings from V1 neurons in a cat. Since dots of various kinds consistently produced responses in neurons located downstream up to V1, in the retina and the LGN, the experimenters presented various dot patterns and looked for responses in V1 neurons. There was occasional firing response in the neurons they recorded from, but it was not clear if the neurons were firing spontaneously or actually responding to the stimuli. After trying out many slides with dot patterns on them, they zeroed in on one particular slide with a dot on it, which seemed to

**Fig. 7.16** An oriented edge (a) and an oriented bar (b)



produce responses with the highest probability. After much further struggle, at a rare “aha” moment in the field of visual neuroscience, the researchers realized that the neuron was responding not to the dot pattern on the slide but to the edge of the slide. The response seemed to occur whenever the lab assistant changed the slide and the slide moved across the visual field of the neuron of interest. Ergo—the neuron is responding to an edge!

Greatly encouraged by this radically new finding, the experimenters prepared slides with various line or bar patterns. Single neurons of V1 were found to respond best when the line pattern was of a particular orientation. The neuron fires away a volley of spikes when a vertical bar was presented; but as the bar was gently rotated, firing dropped rapidly. Larger deviations from the vertical elicited no response at all. In addition to the orientation of the line, V1 neurons also adapted their response to the position of the line. Firstly, neurons respond only when the stimulus lies within their receptive field. When a bar of correct orientation is presented within the receptive field of a neuron, some neurons responded best when the bar is right in the middle of receptive field. Response dropped as the bar is moved away from the center, even though the orientation is kept constant. Hubel and Wiesel named such cells the “simple cells”, in contrast to other cells with more complex response properties which they found later. The ability of V1 neurons to respond to bars of a given orientation is called orientation sensitivity, a property that may be described as a defining property of V1.

Hubel and Wiesel found other neurons which showed orientation sensitivity but their response did not diminish significantly when the bar is moved within their receptive fields. Such neurons were labeled complex neurons. Then, there were neurons that were sensitive to bars of specific length. Response fell when bars that are shorter or longer than the optimal length were presented. In addition to bars, V1 neurons also responded to edges (Fig. 7.16), which are formed between two regions with contrast between them. Then, there are neurons that respond when a bar of a specific orientation moved specifically in one of the two directions perpendicular to itself. If the bar moved in the opposite direction, the response fell. Such neurons are said to be *direction sensitive*.

In summary, we observe a trend in neural responses as we climb up the hierarchy of the visual system from the retina to V1. Neurons of the lower stages look at a world as a mass of dots, perhaps because they break up the visual space into extremely tiny cells. But since the ON/OFF type receptive fields of the lower stages have the ability to enhance contrast, outlines of objects in the retinal images seem to be sharpened.

Therefore, what the V1 neurons get to see are images with sharpened edges, which is perhaps the reason they become specialized at responding to edges and bars. Orientation-sensitive V1 neurons, therefore, seem to be processing the outlines of objects in the visible world. Furthermore, V1 neurons that respond to moving bars, the direction-sensitive neurons, are probably processing outlines of moving objects. On the whole, one could imagine that the retinal image patterns, as they flow through the visual hierarchy, are worked upon systematically, as in an assembly line, with each stage bringing about a specific transformation in the image it received, and passing it onto higher stages for more sophisticated processing.

Since the real world has contours oriented at all sorts of angles, there must be neurons in V1 that respond to all possible orientations, a possibility that raises another question. If there are V1 neurons that respond to the entire range of orientations from 0 to 180°, how are the neurons distributed over the surface of the cortex? To use an analogy, imagine a large class in which the teacher decided to seat the students according to the day of the birth numbered from 1 to 31. Assuming the class has neat rows of seats arranged as a rectangular grid, are the birthdays distributed randomly, or is there a special ordering?

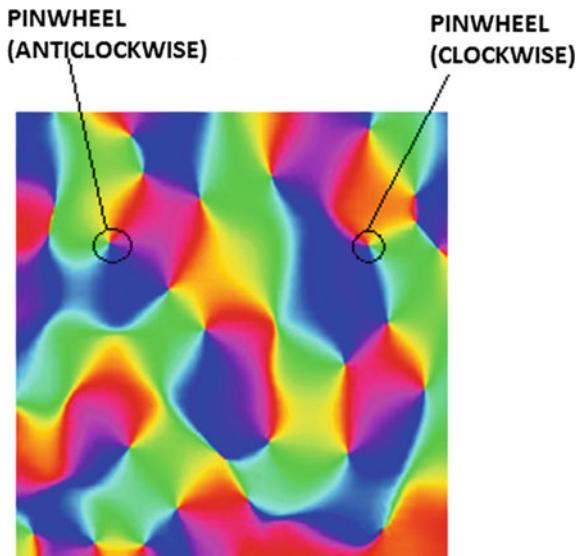
This was the question that was next taken up by Hubel and Wiesel in their systematic studies of the primary visual cortex. First, they studied the pattern of responses of neurons at a single point on the cortex, penetrating all the cortical layers at that point. The cortex is a thin slab of tissue about 2–5 mm thickness, consisting of six layers of neurons. The experimenters penetrated the depth of the cortex with a microelectrode and probed the responses of neurons at different depths, in different layers. It turned out neurons at different depths responded, if they did, to about the same orientation. In some layers, neurons did not show any orientation sensitivity.

Then, the experimenters probed the cortex along the surface. Every time the electrode advanced by 0.05 mm, it was observed that neurons respond to a different orientation. Neurons within a circular area of diameter 0.05 mm respond to the same orientation. Actually, neurons within a cylinder that spans the full depth of the cortex, and has a diameter of about 0.05 mm, respond to nearly the same orientation. These cylindrical units of cortex, labeled *orientation columns*, seem to be some sort of basic building blocks of V1.

Neurons responded to progressively varying orientation as the electrode advanced along a line parallel to the surface of the cortex. A full cycle of orientations is typically visited over a distance of about 1 mm. But sometimes there is a sudden shift in the direction of change in orientation. And sometimes a sudden change of about 90° is observed in orientation sensitivity. It appears that on a short linear stretch, one might observe a neat continuous variation of orientation, but over a larger extent, variation of orientation sensitivity is more complicated.

Mapping orientation sensitivity over the entire two-dimensional surface of the primary visual cortex by probing it with a microelectrode is not practically feasible. An entirely different experimental technique is more suitable for such a measurement. Gary Blasdel and colleagues developed exactly such a technique in 1985. They used voltage-sensitive dyes to image the orientation maps of the visual cortex. Neurons stained with these dyes emit light. Since the dye is voltage sensitive,

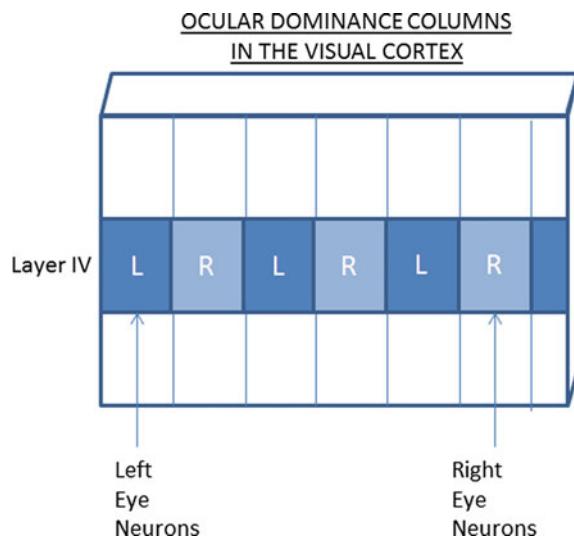
**Fig. 7.17** A schematic of an orientation map with two pinwheels shown



active neurons emit stronger signals, thereby enabling imaging of neural activity. Oriented bar patterns at a range of orientations are presented one after another. The corresponding responses of the visual cortex are recorded and combined using complex image processing techniques, ensuring that neurons that responded to different orientations are labeled appropriately. Figure 7.17 shows a schematic of a typical orientation map observed by 2D imaging methods. The tiny lines indicate the orientation to which the local neuron is sensitive. Note that as you traverse the map in any given direction, the orientation changes gradually completing a full cycle over a distance. What is particularly remarkable is that there are points where neurons have no orientation sensitivity. Around such “orientation-neutral” regions you will find neurons that respond to all possible orientations. If you go around one such orientation-insensitive neuron in, say, anticlockwise direction, you will find orientation sensitivity gradually increasing and completing a full cycle as you return to your starting point on the circle. By analogy to the spokes of a wheel, such points are known as *pinwheels*. Earliest orientation maps obtained from 2D imaging did not have enough resolution to study the structure of the map in the neighborhood of the pinwheels. Imaging at a higher resolution was performed more recently by R. Clay Reid and colleagues. Observations made at lesser resolution were reconfirmed by the new findings. Around every pinwheel, the experiments showed an orderly arrangement of the full cycle of orientations.

Orientation maps are yet another example of the strategy of fine-grained division of labor that underlies the functional architecture of the visual system. Like the fellow who was good at turning a certain nut in the earliest assembly lines, each neuron in V1 appears to be good at just one microscopic function: ability to sense a tiny line segment with a certain orientation placed at a certain position in the visual field.

**Fig. 7.18** A schematic of the ocular dominance columns in the visual cortex



Neurons with such specialized sensitivities to orientation are distributed in intricate patterns over V1.

Neurons of V1 respond to another visual property in addition to orientation. Remember that nerve fibers from the two eyes are routed such that information about one half of the visual field is projected to the opposite half of the brain. To enable that, each eye sends out fibers from a half of its retina to the appropriate side of the brain. Therefore, each of the visual brain receives input from both eyes. A question arises now: does a given V1 neuron respond exclusively to inputs coming from only one eye (right or left) or from both the eyes?

Like in the case of orientation sensitivity, there is a division of labor in V1 in terms of response to signals coming from the two eyes. Some neurons respond exclusively to inputs from right eye, while others respond to the left eye signals. Since, in the responses of these neurons, one eye dominates the response of the other, this property is known as ocular dominance. There are roughly equal number of “right eye neurons” and “left eye neurons.” These two populations of neurons are arranged in intricate patterns forming what is referred to as ocular dominance map. Construction of an ocular dominance map is relatively easier compared to the elaborate procedure followed to construct orientation maps, where response to a range of orientations needs to be summated. In one of the earliest attempts, Wiesel and Hubel injected radioactive tracers into one of the eyes of a rat. The tracer winds its way up the optic nerve, crosses the synapse at the LGN, and arrives at the V1. Neurons that respond to the eye in which the tracer was injected, pick up the tracer, and therefore emit radiation, which can be used for imaging, while the “bright” neurons in the image correspond to the eye which received the tracer, the rest of the neurons correspond to the other eye. Image obtained using such methods revealed exquisite band patterns bearing strong resemblance to zebra stripes (Fig. 7.18).

In addition to orientation sensitivity and ocular dominance, V1 neurons respond to another important property of visual stimuli—color. Amidst orientation columns, there are barrel-shaped clusters of neurons fondly named *blobs*! Neurons inside blob regions are particularly sensitive to color but do not respond to orientation.

Naturally, tangles of lines and blotches of color are not neatly segregated in separate regions of the visual field; they are inextricably mixed everywhere. In V1 too neurons that respond to orientation, ocular dominance, and color are grouped together into small regions that were named hypercolumns by Hubel and Wiesel. A single hypercolumn, about  $1 \text{ mm}^2$  in area, contains orientation columns of full range of orientations, ocular dominant neurons of both left and right variety, and blobs that respond to a range of colors. It, in fact, seems to be a natural way of organizing the cellular level functional units of V1. To use a gastronomical metaphor, imagine how the tables might be arranged at a large banquet where a large number of guests are expected. The menu, broadly divided into appetizers, main course, and deserts, is set on three adjacent tables. Copies of such triple tables are repeatedly arranged over a large area so that a guest can walk up to the nearest triple table, covering the gastronomical distances that separate him/her from the feast. Thus, a hypercolumn is a basic, composite, visual processing unit, a tile in the mosaic of the primary visual cortex.

The three different types of V1 neurons that we have encountered so far represent key functional units of information processing in V1. They begin to channelize the streaming visual information into semi-independent dimensions of form, motion, and color. The orientation-sensitive neurons process outlines of objects. The direction-sensitive cells process motion. Neurons with ocular dominance form part of the neural machinery that can extract three-dimensional form. Blob neurons respond to color. This segregation of visual dimensions, which has actually begun at the retinal stage, progressively develops as we climb up, via LGN, to V1 and beyond, consummating in a sense, in specialized visual cortical areas dedicated to different aspects of visual processing—color, form, motion, etc.

## Visual Maps and Cortical Blindness

While it is important to describe the extremely fine-grained fashion in which functions are distributed over individual cells in V1, it is equally important to mention in this context that the responses of individual V1 neurons are highly malleable and are not stuck in a state of deep freeze. This malleability is in fact one of the defining properties, not just of V1, but of the entire brain. It is this ability to change, adapt to changing environments, and pass on the results of such adaptation to progeny through language and other media of learning that has rapidly propelled us from the state of cave-dwelling hunters and gatherers to one that can boast of a complex urbanized civilization.

Neuroscientists found that the brain is particularly changeable at certain vulnerable periods in its development. This vulnerable period, popularly known as the

*critical period*, is different for different functions of the brain. Every parent experiences with legitimate pride the impressive explosion in his/her child's vocabulary that occurs suddenly when the child is around 2 years old and progresses rapidly over the next few years. Critical periods have been found to exist in vision too. In fact, there are separate critical periods for different properties of vision. Critical periods have been studied in great depth in case of a variety of animals by experimentally creating visual deficits at various stages in the animal's life and marking out the period where deficits have the most catastrophic consequences. Such experiments are obviously impermissible in humans. But critical periods in vision are observed in humans too in cases where they were born with a visual impairment, or when the impairment happened to occur during the critical period with devastating consequences.

The earliest experiments on critical periods in vision were performed by the pioneers Hubel and Wiesel again by inducing monocular deprivation in cats. One of the eyelids of the animal is sutured, thereby rendering the corresponding eye temporarily blind, while the other eye is left intact. After a stipulated period, the sutures are removed and the affected eye is allowed to respond to light. The most drastic effects of monocular deprivation, albeit temporary, were observed in the ocular dominance patterns of V1. Note that neurons that respond to left and right eye inputs are still adequately segregated in distinct layers of LGN. However, in V1, within an area of about 1 mm in diameter, neurons that respond to both eyes are found. When one of the eyes is sutured, neurons that were earlier responding to the unsutured neuron fell silent. A good fraction of these neurons, now unemployed, gradually tune themselves to respond to the stimuli that arrive from the open or intact eye. (We may recall a very similar situation that we encountered in Chap. 6 where dramatic changes in the somatosensory map of a monkey were observed when one of its fingers was amputated. Neurons in the map that earlier responded to the missing finger, found gainful employment in responding to fingers adjacent to the missing finger.) After a certain period, whose length is the shortest during the critical period, there are hardly any neurons left in the V1 to respond to the stimulus coming from the recently sutured eye. The black and white zebra stripe patterns of ocular dominance maps of Fig. 7.19, with black and white patches of nearly equal area, now become completely skewed. The map is now either nearly completely black or white depending on which eye was deprived. Since V1 neurons do not respond to the stimuli from a certain eye that eye is practically blind. The results are most dramatic when the monocular deprivation occurs during the critical period. In cats, whose critical period occurs around the age of 4–5 weeks, 1 or 2 days of deprivation is sufficient to produce lasting changes in the animal's ocular dominance maps, rendering the animal permanently blind in one eye.

Experiments on ocular deprivation show the phenomenon of blindness, or its positive opposite-vision, in new light. According to lay understanding, successful vision depends almost exclusively on intact eyes. It is generally thought that seeing occurs, as if by magic, when the eyes capture the image on the retina and transmit the same to the brain. But the above experiments demonstrate the dependence of response properties of the visual cortex on the visual stimulations of the world. Therefore, there are two preconditions for successful seeing: one on the side of

**Fig. 7.19** A schematic of ocular dominance map



machinery for transducing light, namely, intact eyes; the other on the side of the brain, namely, well-developed maps in the visual cortex.

Our understanding of the many subtle and enigmatic aspects of blindness has grown thanks to the study of case histories of human subjects with visual impairment. Typically, these were cases of individuals who were blind, from their childhood through their adulthood due to an impairment of their eyes. Although the function of the eyes was surgically restored at some point in the adulthood, these individuals failed to develop full visual capabilities of one with intact vision.

One of the earliest reported case study of this kind dates back to 1728 when an English surgeon named William Cheselden restored vision by removing cataract in a congenitally blind 13-year-old boy. Since the retina was intact, the boy had functioning eyes after the operation. Despite this positive development, the boy experienced serious difficulties in seeing. He had no idea of space or distance. Two-dimensional drawings and pictures confused them since he did not know how to relate them to the three-dimensional reality. He was rather slow in understanding and deciphering the visual world, perhaps since whatever little sense he could make of it depended on his efforts to cross-validate and substantiate it with his sense of touch.

A similar and more dramatic case, widely publicized through the writings of the eminent neurologist Oliver Sacks, was found only a couple of decades ago. In 1991, Sacks happened to see a 50-year-old patient named Virgil, who was blind since birth again due to cataract. At an age of 45, on the insistence of his fiancé Amy, Virgil allowed himself to be operated for cataract. Once the cataract was removed, since the retinas were intact, Virgil began to see. After the initial euphoria about the restoration of sight after such a long period of darkness, Virgil family began to notice Virgil's difficulties in coping with the newfound capacity. It was as though Virgil was struggling to make sense of what he is seeing. In Amy's words, he was "like [a] baby just learning to see, everything new, exciting, scary, unsure of what seeing means."

Post surgery what Virgil was able to see was color, form, and movement; what he could not see was meaning. He had trouble fixing his attention on an object in order to comprehend. He would make an abortive attempt to fix attention on a target before

his attention drifts away, only to return after some time, repeating the exasperating drama of trying to know a thing. A close inspection of the retina of the affected eye revealed degeneration of macula, the central part of the retina. Macula has a high density of photoreceptors on which images of fixated targets fell. Its degradation explained why Virgil had trouble with fixing vision on targets.

Virgil was able to see motion and in fact was fond of watching moving objects. He would feast on sights of large colorful vehicles like the bright yellow school buses, or blood red SUVs. Bright neon lights on billboards appealed to him. But he did not like it if the scene is too cluttered, had too many things in it. Therefore, he could rarely read a complete word; he would read a couple of letters and tried to guess the entire word. The crowding of letters in a long word seemed to bother him. This discomfiture clearly showed up in one of his visits to a grocery store. The shelves, the vegetables, the aisles, and the people—all packed in a single vista intimidated him so much so that he insisted on leaving the store immediately. At the other end of the spectrum, simple uncluttered views, like that of rolling green hills surrounded by vast verdant fields brought him joy.

As Virgil's struggles in coping with the visual world continued, an unfortunate accident resulted in a bitter turn of events, dashing the hopes of ultimately restoring full vision. One of his lungs was dangerously shrunken due to an attack of lobar pneumonia. The accident led to breathing problems and insufficient oxygen in the blood. Low oxygen levels in the blood affected the function of his retinas. His vision began to fluctuate painfully—he could see one moment and was blind the next. The crisis only worsened with time. Virgil was forced to return to his world of darkness, with which he was accustomed to all his life, after a brief exposure to the joys of seeing light. The surgery might have opened the doors of the eyes, but due to prolonged disuse, the doors of the visual cortical areas have been closed. And it is these stubborn cerebral doors that refuse to open or open fully, turning the initial joy of seeing light into the trauma of inability to make sense of the world.

Virgil's story demonstrates that there is a lot more to vision than the simple, camera-like function of the eyes. It drives home, albeit with some bitterness, the myriad subtle and as yet poorly understood conditions under which the experience of seeing occurs.

## V2

Let us now return to our narrative of the sequence of visual transformations that occur as we ascend the visual hierarchy from the retina to an array of visual cortical areas. Just as the primary visual cortex was named V1, higher visual cortical areas were named after a simple nomenclature as V2, V3, etc. The secondary visual cortex or V2 is adjacent to V1, in a sense circumscribing V1, while it has been possible to attribute to the neurons of V1 the simple function of response to oriented features, for a long time after the pioneering studies of Hubel and Wiesel, the function of V2 neurons remained unclear. Being next to V1 in hierarchy, it was expected that V2

**Fig. 7.20** The stimuli that produced the strongest responses in V2 neurons



has important visual functions. Monkeys with V2 lesions suffered from the inability to perform complex spatial tasks.

In order to investigate the functions of V2 neurons, Jay Hegde and David van Essen presented a variety of visual patterns, more complex than the simple oriented bars that elicited response in V1. The patterns included complex gratings of parallel stripes, with a range of orientations and spacing; black and white spiral patterns; crosses and stars; and angles and arcs. The responses of V2 neurons to these patterns were assessed. It was discovered that patterns that elicited strongest responses are angles and arcs (Fig. 7.20).

It is possible to present simple, geometric arguments to account for why V2 neurons should respond best to angles and arcs. V1 neurons look at short segments of contours of objects in real-world images. Since short segments of curves can be approximated by straight lines, we can account for the orientation sensitivity of V1 neurons. V2 is at a higher level in visual hierarchy than V1. A single V2 neuron receives inputs from a large number of V1 neurons, and therefore has a large receptive field than that of V1 neurons. V2 neurons, therefore, are capable of recognizing longer segments of object contours, which are curvilinear features. We can similarly explain the response of V2 neurons to angles. Consider a V2 neuron which receives inputs from two V1 neurons that respond to orientations  $O_1$  and  $O_2$ , respectively, should be capable of responding to an angle pattern that is constructed of the two orientations (Fig. 7.20a).

Minami Ito and Hidehiko Komatsu had taken the above line of work further and studied the spatial distribution of angle sensitivity in V2. Similar to orientation maps of V1, they found a map of angles in V2. Neurons that are close to each other in the map tend to respond to similar angles. Although not all neurons responded to angles, among those that responded, nearly equal number of neurons responded to sharp angles and wide angles. The corner of the angle typically coincided with the center of the receptive field of the neuron. An interesting feature of the Ito and Komatsu study was that neurons that responded to an angle, often also responded to straight lines that corresponded to the arms of the angle. A given neuron responded to either a single arm or both the arms of the angle. Such a response pattern lends weight to the idea that the responses of V2 neurons are constructed of the orientation-sensitive responses of the V1 neurons.

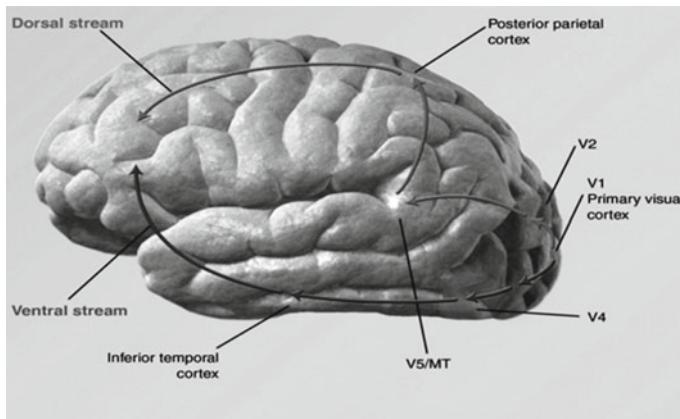
## Perceiving Movement

So far, our discussion of the function of the visual system has been confined to perception of static images, delineating outlines as a step toward identifying visual objects. Our experience of the visual world, however, does not consist of a series of snapshots of the world frozen in time. The world we see is one of a constant flow of life, of incessant change. From striking large-scale movements like the crashing waves, or a lightning strike, to subtler ones like the trembling lips or a flushing face, movement is the soul of the world. It is hard, if not impossible, to imagine a world without motion, immobilized for all eternity. But, ironically, our intuitive understanding of the world has its roots in brain's (a healthy brain's, that is) inner workings. Any jamming of our neural machinery brings about disconcerting alterations in our experience of the world.

Such jamming of neural machinery responsible for perceiving motion was perhaps first observed in 1978 when a 43-year-old woman who complained of severe headaches was admitted to a hospital in Munich. She was discharged after a few days. She suffered from a stroke in a region that lay on the border between temporal and occipital lobes. Nineteen months later she was examined closely by neurologist Joseph Zihl who found that the patient was nearly normal in most respects but suffered from a disorder marked "loss of movement perception in all three dimensions." She could not perceive the flowing, continuous aspect of motion and saw movement as a series of disconnected snapshots, like the scenes from a discotheque. The patient, usually referred by her initials as LM in neuroscience literature, had, for example, trouble of pouring tea into a cup since the liquid appeared to be "frozen like a glacier." She had a similar difficulty in crossing a busy road since she could not perceive a vehicle gradually approaching her, because one moment the object was there far away, and the next moment it is right here. It is as though the whole world of movement fell apart. Visual neuroscientist Semir Zeki labeled this condition *akinetopsia* or inability to perceive motion.

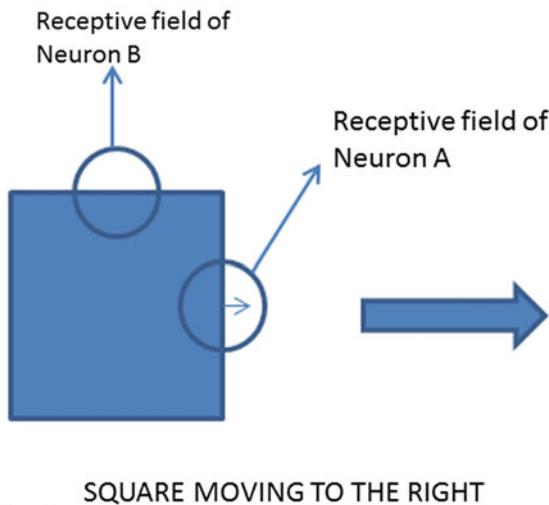
Subsequent studies in both humans and monkeys were able to precisely identify a cortical region that is specifically responsible for motion perception. This region known as the Middle Temporal (MT) area (Fig. 7.21) is actually in the whereabouts of the region that was affected by stroke in LM's brain. Neurons in MT were found to respond to moving spots or bars, and therefore thought to offer the neural machinery to process motion. But we have seen similar neurons in V1 too. The direction-sensitive neurons of V1 indeed respond to bars moving in a specific direction perpendicular to themselves. How is the response of MT neurons different?

MT neurons sense motion of the entire object, whereas the direction-sensitive neurons sense the direction of an oriented line, or an edge which might be part of a moving object. To understand the difference between sensing object motion as opposed to motion of a small line segment, consider the square moving to the right in Fig. 7.22. Imagine a direction-sensitive V1 neuron A whose receptive field overlaps with a part of the right edge of the square. Assuming A is tuned to vertical bars moving toward right, we would expect to fire in response to the moving square.



**Fig. 7.21** Figure indicating the location of V5 or MT

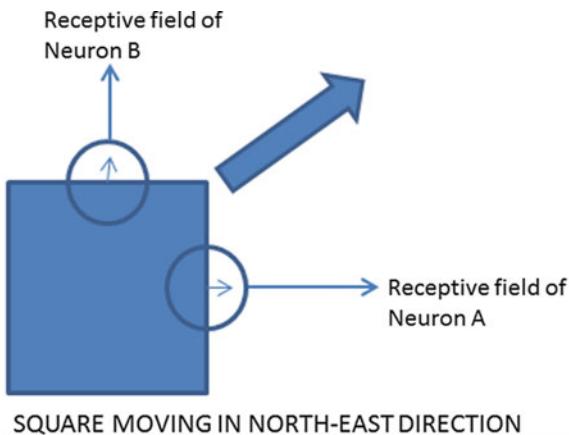
**Fig. 7.22** An illustration of the aperture problem in which the motion of one of the edges is same as that of the whole object



Now, consider another direction-sensitive V1 neuron B whose receptive field lies on the upper edge of the square. Since there is no relative motion of the upper edge of the square, neuron B fails to detect any motion. This simple illustration demonstrates that responses of direction-sensitive neurons of the kind found in V1 are incapable of reliably estimating the whole body motion. In the above case, while A sensed the object motion correctly, B failed to sense anything.

The situation can get even more complicated, as shown in the illustration below. Figure 7.23 shows another square moving in the northeast direction. A and B are direction-sensitive V1 neurons tuned to rightward moving vertical bars and upward moving horizontal bars, respectively. In this case, both the neurons fire. But, whereas

**Fig. 7.23** An illustration of the aperture problem in which the motion of the edges is different from that of the whole object



A signals that the object is moving rightward, B thinks it is moving upward. Both A and B have something to say about the real motion of the square. The problem of finding the real motion of the object from the partial evidences supplied by the activities of neurons (like A and B) looking at motion of edge segments, is known as the *aperture problem*. It is likely that the aperture problem is solved by some higher area in the visual system which receives inputs from V1, directly or indirectly. MT appears to be one such an area.

Tony Movshon and his colleagues have explored how the neurons of MT solve the aperture problem. In these studies, the choice of moving objects required special consideration. The objective of the study is to monitor activities of both V1 and MT neurons and understand their differences. Since response to orientation is the common factor that links V1 and MT neurons, use of whole objects like, for example, a square poses a special challenge. A V1 or MT neuron that responds to a part of the edge of the square at one position might cease to respond the moment the square moves forward since its edge may no longer lie on the receptive field of the said neuron. Therefore, it becomes imperative to use patterns made of parallel lines. Furthermore, since the study must bring out the contrast between responses to motion of an edge and responses to whole body motion, the pattern must involve at least two sets of parallel lines. Therefore, the researchers used plaid patterns, made of sets of parallel lines, intersecting at right angles. Moving plaid patterns were presented to monkeys while the responses of V1 and MT neurons were monitored. The recordings clearly revealed the difference between the response of V1 and MT neurons. Most V1 neurons, and some MT neurons too, responded best to the movement of components of the plaid—the sets of parallel lines. They did not respond to the motion of the whole plaid. But a small fraction of the MT neurons responded to the direction of the whole plaid. Thus, it appears that the visual system solves the aperture problem by stages—motion of single components is detected at the stage of V1, while more large-scale motion is detected at MT.

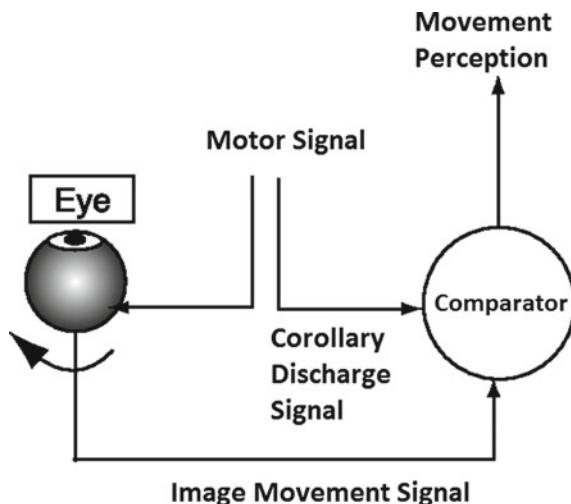
But the activity of MT neurons is not the final word on motion perception in the brain. Brain's motion perception has to deal with a more fundamental aspect of motion, which we have not brought up till now in our discussion. Physicists since the time of Galileo have been aware that all motion is relative. Relativity of motion is something brain cannot escape from, and had actually evolved excellent mechanisms to cope with it. Until now in our discussion, we have been considering a simplified depiction of motion in which a moving object in the real world produces a moving retinal image, which is processed by the visual system. But an object moving in the world need not always produce a moving retinal image. When you see trees and telegraphic poles whizzing past while you sit quietly in a train's compartment, you know obviously that the motion is due not to the trees themselves, but due to the train that is carrying you. When your eyes are closely tracking a cricket ball soaring into the skies, propelled by Mahendra Singh Dhoni's fierce strokes, the retinal image of the ball may not be moving much, but your brain is not fooled into thinking that the ball is not moving. In case of the ball, the motion of the eyes in some sense is canceling the motion of the image of the ball on the retina, but the brain seems to factor in the motion of the eyes to recognize the true motion of the ball. As you walk toward a painting hung on the wall, you see the painting grow in size in front of your eyes. But you know that it is your relative motion that is causing the expansion, since your brain factors the feedback from your apparatus of locomotion into its visual perception. As you race along the tortuous tracks of a roller coaster and see the world spiraling insanely in front of you, you know well that the culprit is really your very own ego hurtling out of control through the 3D world.

Therefore, for the brain to make accurate judgments of true motion in the world, it must consult two sources of perceived motion—the changing retinal images on one side and the motor commands that generate movements in ourselves, in our own body relative to the world, that are responsible to perception of motion. Brain combines the knowledge of self-generated bodily movements with retinal images and arrives at a more reliable understanding of the true motion in the world. There are several types or rather several levels of bodily movements that are involved in motion perception.

At the simplest level, there are these darting movements of the eyes, known as *saccades*, which can produce moving retinal images when the object out there is actually still. The opposite situation occurs when the eyes are actually smoothly tracking a moving object in such a way that its retinal image is quite still but motion is perceived by the brain. In fact, an object that suddenly moves in front of you elicits a tracking movement of the eyes as a reflex action, a reflex known as optokinetic reflex that develops in infancy. The next level of bodily movement consists of movements of the head. Like the eyes, movement of the head, with the eyes moving along with the head, can produce moving retinal images. Sometimes the eyes move in a direction opposite to that of the head so as to stabilize the retinal image. Such corrective movements occur in what is known as Vestibulo-Ocular Reflex (VOR) in which sudden movements of the head elicit corrective movements of the eyes in order to stabilize retinal images.

In the middle of the last century, neurobiologist Robert Sperry proposed the *corollary discharge theory* to explain brain's ability to factor in internally generated move-

**Fig. 7.24** A schematic of the corollary discharge theory



ments in accurately recognizing the true changes in the world. Figure 7.24 depicts using a simple block diagram the essence of corollary discharge theory. The system involves three signals—(1) the Image Movement Signal (IMS), (2) the Motor Signal (MS), and (3) the Corollary Discharge Signal (CDS). The IMS originates in the retinas and climbs up the visual hierarchy to higher visual cortical areas. The MS is the command sent to the eyes to move. The CDS is a parallel copy of the motor command, sometimes referred to as the *reafference copy*. The CDS is sent to the visual cortical areas where the two sources of information related to movement perception (self-motion and retinal images) are combined. The CDS is used by the brain to predict the most likely change in the retinal image produced by the movement of the eyes just commanded by the brain. The visual cortical region concerned, compares this predicted scene with the actual scene from the retinas. If the two scenes match perfectly, it means that the perceived motion is caused solely by the movement of eyes, and there is no real motion out there. If the two scenes do not match, the brain infers that there is some real motion occurring in the world out there (Fig. 7.24).

When objects moving against a fixed background of the world are perceived, their retinal images also exhibit local changes corresponding to the moving objects. On the contrary, motion perceived in the world due to self-motion relative to the world is of a more global nature consisting of specific kinds of patterns referred to as optic flow. American psychologist James Gibson first introduced the concept of *optic flow*. Gibson was a proponent of the ecological approach to psychology which deviated from traditional approaches that studied animal behavior in conditioned laboratory settings. The ecological approach argued for a need to study animal behavior in their natural habitats, in their interactions with nature. In this process, Gibson recognized the importance of optic flow for an animal in making it aware of its self-motion with respect to the world, and use that awareness to plan its locomotion. Optic flow often comes in certain standard patterns. For example, movement of the observer

toward a target makes the target appear to expand. Similarly, movement away from the targets produces an appearance of contraction. An approaching motion of an observer toward a target, simultaneously rotating about an axis passing through the observer and the target, produces the appearance of an expanding spiral. A similar motion of a withdrawing observer produces a contracting spiral.

There is a visual cortical area known as Middle Superior Temporal (MST) area adjacent to the MT area, where neurons were found to respond to optic flow. Neurons that respond specifically to expansions and contractions of the visual field corresponding to translational motion, and to spiraling patterns corresponding to rotational motion have been discovered in MST. Some MST neurons have been found to recognize the real motion of external objects: they respond when the object moves but not when only the eyes move. Furthermore, subjects with damaged MST area felt an illusory movement of the world whenever their eyes moved.

## Recognizing Complex Objects

All visual processing in the brain seems to be answering one of two questions about objects in the visual world—"what is it?" and "where is it?" The first question refers to recognizing and identifying an object based on an understanding of the multitude of its properties (color, motion, etc). The second question places the objects against the backdrop of the surrounding space, and other objects, and seeks to understand the relationship of the object to its milieu. (Is it to the right or the left? Is it behind another object or in the front?) As we ascended the hierarchy of visual cortical areas, from V1 and beyond, we noticed that the lower cortical areas performed preliminary analysis and extracted information like outline, movement direction, depth, or color. This sequence of visual cortical areas culminates in a cortical area called the Inferotemporal (IT) area. Unlike the neurons of the lower layers, neurons in IT area respond to the presence of complex objects in the visual field. The IT neurons gather the more preliminary information from below, and use it to recognize complex objects.

A beginning of our understanding of the link between temporal lobe areas and visual processing can be traced back to the radical experiments of Heinrich Kluver and Paul Bucy who studied the effects of temporal lobe removal in monkeys (see the description of Kluver–Bucy syndrome in Chap. 10). The animals lost the ability to recognize objects visually and displayed an exaggerated oral behavior—they began to put objects in their mouth indiscriminately in a putative attempt to understand them. In addition to these impairments of sensory function, the animals also showed altered emotional behavior. They became unusually tame and exhibited abnormal sexual behavior. Subsequent studies teased apart the sources of the two kinds of impairment observed in Kluver–Bucy syndrome. It turns out that the emotional impairment had its origins in the loss of amygdala, a key temporal lobe structure, while the deficits in visual discrimination had its origins in the inferotemporal area. Later studies demonstrated that bilateral lesions in the IT area caused impairments in visual object

discrimination. Such impairment was observed even when the neural activity of the IT area was temporarily disrupted using electrical stimulation.

The involvement of IT area in object recognition is understandable considering the large receptive fields of the IT neurons. We know that receptive field sizes grow as we ascend the visual hierarchy. The receptive field size of retinal ganglion neurons is only  $1^\circ$ , which increases to  $4^\circ$  in V4. The size increases dramatically to  $16^\circ$  in posterior IT area and to  $150^\circ$  in anterior IT. Large receptive fields are essential for neurons to respond to spatially extended complex visual objects.

The idea of neurons responding to complex visual objects is intuitively appealing and seems to carry a natural explanatory power. Since tuned or specialized responses of neurons (see Chap. 6) is a ubiquitous property in the brain, it is quite compelling to carry over the idea from simple visual primitives like oriented bars all the way to complex shapes like faces and spanners. But responding to orientation is very different from responding to complex objects. The orientation space spans a finite range of angles ( $0$ – $180^\circ$ ), assuming a discreteness in representation. Therefore, it is possible, and is even meaningful, to construct a map of orientation. But how do we map the space of complex objects? A map implies knowledge of the dimensions of the underlying space. What are the basic parameters of the entire space of complex visual objects? Is there a *visual alphabet* using which any visual object of the world, known and unknown, can be assembled? Without the knowledge of some sort of underlying parameters of visual object space, one will have to have separate neurons uniquely respond to every single object in the world, which leads to an explosion of the number of visual object recognizing neurons.

In order to address this deep difficulty, Keiji Tanaka and colleagues began with a simple question: When a neuron responds to a complex object like, say, a chair, what exactly is it responding to? What is it about a chair, wherein lies its “chairness”? that elicits a response from the said neuron? Does it need to have four legs or would three suffice? Does it need to have armrests or are they exempt? Does the neuron respond to one specific type of chair or is it liberal enough to respond to a large class of chairs? Can it be deceived by the chair-like appearance of a four-legged animal?

Tanaka and coworkers set out to find the essence of the complex patterns that evoked responses in IT neurons. Once a neuron that responded to a specific complex pattern is found, the researchers systematically began to simplify the complex pattern until the neuron that earlier responded to it did not anymore. Consider the picture of a screwdriver with shading to give a rich three-dimensional effect (Fig. 7.25, leftmost). A first step to simplification is to remove the shading making it look like a flat picture of a screwdriver. Neural response to this transformed image is still significant. In the next incarnation, the screwdriver is simply an ellipse with a line protruding out. Neural response is actually the maximum at this point. But the moment the protruding line is removed, leaving only the ellipse, neural response falls to nearly nothing (Fig. 7.25). Note that the above-described process of pattern reduction is done on simply intuitive lines and is not governed by any formal process of mathematical simplification.

A large number of real-world complex patterns are similarly reduced. The resulting patterns are definitely more complex than oriented lines that are V1’s favorites,



**Fig. 7.25** Progressive simplification of natural visual stimuli used in Tanaka's experiments

but less complex than the real-world objects from which they were derived. These patterns seem to be some sort of archetypes of the visual world by which the visual brain breaks up and represents visual objects. A few logical questions are in order: Is there a map of complex shapes in IT? Is there columnar structure? Analogous to orientation columns in V1, the researchers found columnar structure in IT. Neurons at various depths at the same point, or nearby neurons within a diameter of about 400  $\mu\text{m}$ , responded to similar objects.

Further exploration of the responses of IT neurons to complex visual objects revealed other elegant aspects of how real-world objects are represented in this part of the brain. Until now, we have been depicting visual objects as complex *two-dimensional* patterns. But real-world objects are three dimensional. They must be associated with not one but a whole continuum of two-dimensional patterns, where each pattern corresponds to a different view of the same object. Therefore, a neuron that claims to recognize a three-dimensional object must respond not just a single view of the object, but several, if not all.

The appearance of an object changes as our spatial relationship with the object varies. Objects seem to grow larger as we move closer to them. Change in appearance is even more complicated as we move around an object sweeping a large angle. But if the angle swept is relatively small, the object appears to shrink or expand in the direction of our motion. Change in distance changes the size of the appearance of the object. On the other hand, if our motion is in horizontal direction, then the shrinkage or expansion also occurs in horizontal direction, leaving the height invariant. In other words, such a motion of the observer changes the aspect ratio of the appearance of the object. An ability to recognize three-dimensional objects implies a certain level of invariance in recognition with respect to changes in scale and aspect ratio. Such invariance was indeed discovered in IT neuronal responses by Hosein Esteky and Keiji Tanaka. When two-dimensional patterns were presented at different sizes, neurons were found to respond significantly to a range of scales spanning over two octaves. Similarly, when planar patterns were presented at different aspect ratios, neurons were found to respond with greater than 50% of the highest response even over a variation of three octaves in aspect ratio.

Just as other visual maps encountered so far in our discussion, the “object maps” of IT area are not cast in stone and can be altered by changing stimulus conditions. In order to study the effect of training on IT neuronal responses, Tanaka and colleagues trained monkeys to recognize simple, artificial, and geometric patterns by giving the animals juice rewards whenever they recognized the patterns accurately. The training went on for 3 to 4 months. After training several monkeys on such geomet-

ric patterns, the researchers compared the neural responses of the trained monkeys with those of untrained monkeys. All recordings were made from the IT areas. The researchers considered the highest response of individual neurons to trained geometric patterns, with the responses to other natural untrained objects. Responses to the trained geometric patterns are much higher in case of trained animals compared to the same responses in untrained animals. These studies showed that the responses of IT neurons can be altered by stimulus training.

While the IT area processes complex objects in general, without preference to any particular family of objects, within the temporal lobe, not too far from the IT area there is a separate cortical area that specializes in recognizing faces. At a first brush, it might seem surprising why the brain allocates an entire area to represent one very specialized class of objects. But then again considering that brain's responses are modulated by the importance or relevance of the stimulus, it is understandable why brain treats faces with utmost importance. Even in the present century, the most important entities that we end up dealing with in quotidian life, entities that matter the most for our welfare and survival on this planet, are flesh-and-blood humans displaying real faces, and not some sleek-bodied electronic devices with colorful displays and internet connection.

The possibility of existence of a separate brain area for face recognition first emerged from clinical studies of German neurologist Joachim Bodamer in 1947. One of the cases described by Bodamer included a 24-year-old man who suffered from an inability to recognize faces due to a bullet wound to his brain. The patient was able to recognize non-face objects. He could, however, identify people by the sound of their voice, their touch, or gait. Bodamer named this deficit prosopagnosia, a word derived from Greek words *prosapion* = face and *agnosia* = lack of knowledge.

Neurons that specialize in face recognition were found in an area called fusiform area located underneath the temporal lobe not too far from the IT area. The stimuli that produce strongest responses in these neurons are faces, real, or fabricated, like plastic models or video displays. They did not respond to other primitive visual features like gratings, oriented bars of patches of color, and features that elicited characteristic response in other visual areas. The neurons also displayed some level of scale and position invariance in their recognition of faces. They responded equally well when faces are scaled up and down or moved around in their receptive fields. What is it about that face that triggers response in these neurons? Do they regard the face as a particular assemblage of facial features (nose, eyes, etc), or as a random distribution of facial components? In order to test this question, researchers presented fabricated face-like, chimerical images in which the eyes, nose, and mouth are at unnatural positions. Such images, however, failed to produce any significant response from the face recognition neurons which seem to be sensitive to the relative positions of the components of a face. They were found to be particularly sensitive to the distance between the eyes, distance between the eyes and the mouth, and the manner in which the hair overlies the forehead. When parts of the face are removed in the display, responses fell accordingly but did not disappear. Responses to cartoon figures of face were weak. These studies demonstrated that the neurons of fusiform area are truly specialized for this specific class of patterns called faces.

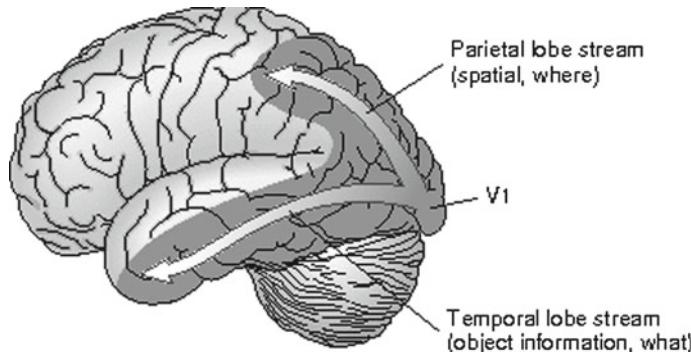
Existence of an exclusive class of neurons for a specialized class of patterns like faces brings up a crucial problem of representation. Is there a one-to-one mapping between faces and neurons of fusiform area? Are there neurons that are exclusively allotted to processing individual faces? Such a notion is exemplified by the hypothetical *grandmother cell* which is thought to respond exclusively to the face of one's grandmother. Or are single faces represented by a distributed network of neurons? The question was settled in favor of the second option by single-cell recordings from macaques observing face images. It turned out that response of single neurons was not completely predictive of specific face images. In other words, responses of a network of neurons must be consulted in order to identify the face that is being viewed. Such *distributed* encodings of external stimuli are obviously more robust. In case of single neuronal encoding, loss of the neuron would completely eliminate the brain's ability to respond the stimulus. Such a crisis can be avoided if the burden of representation is spread out over a population of neurons.

## Pathways of Knowing and Doing

In the preceding sections, we have visited several visual cortical areas with their distinctive functions. We saw a progressive analysis of the moving visual stimulus from the retina to the V1 and beyond. We described motion processing in MT and MST areas. We considered how more complex objects are comprehended in the IT area. There are other key areas which were not dwelt upon in this brief account of the visual system.

There are visual cortical areas predominantly involved in color processing. Damage to certain visual cortical areas leads to a syndrome called cerebral *achromatopsia*, an inability to perceive color. These people experience the visual world in shades of gray and do not perceive any color. It must be distinguished from other color-related impairments like, for example, *color anomia* which refers to an inability to name colors. Both functional magnetic resonance imaging of normal subjects and post-mortem studies on achromatopsic patients revealed that lingual and fusiform gyri of the occipital lobe are involved in color processing.

Then, there are cortical areas involved in perception of depth and 3D shape of objects, which is one of the key functions of the visual system. Stereovision is a crucial mechanism used by the visual system to perceive depth, though it is not the only mechanism. Several other cues—like occlusion or shading—also supply cues to interpret depth. The retinal images of the two eyes are similar but not identical. The visual system uses this difference or disparity between the two retinal images to derive depth information. A key problem that underlies stereovision consists of relating the parts of the two retinal images? How does the brain know that the two tree images seen, one in each of the retinal images, correspond to the same external tree? The need to map points one retinal image to the points on the other is known as the *correspondence problem*. Depth is assessed with respect to a fixation plane, the plane on which the lines of sight of the two eyes meet. Brain interprets the two retinal



**Fig. 7.26** The “where” and “what” pathways

images of a point on the fixation plane as having no or zero disparity. Points that lie beyond or to the fore of the fixation plane produce disparity. There are neurons in the visual system that respond to each of the three types of disparity. The “far” cells, “near” cells, and “tuned zero” cells respond to points beyond, before, and on the fixation plane, respectively. Disparity sensitive cells were found in several visual cortical areas including V1, V2, MT, MST, and IT areas.

Although we have been describing the functions of various visual cortical areas as if they function in isolation, one must remember, like with everything else in the brain, that these areas work as part of a network, occupying their unique places in the complex visual hierarchy. Visual neuroscientists have identified two prominent pathways beyond the primary visual cortex—one leading into the temporal lobe terminating in the IT area and the other leading into the posterior parietal cortex. We have earlier remarked that all visual processing seeks to answer two important questions about the visual world: “what is it” and “where is it?” The first question refers to identifying objects based on their attributes. The second refers to positioning of the object in its spatial context. The first question seems to be addressed by a visual pathway that leads to the IT area from the V1, a pathway that is, therefore, called the “what” pathway, or the “ventral” pathway. The second question is addressed by the pathway that leads from the V1 to the posterior parietal cortex, a pathway labeled the “where” pathway or the “dorsal” pathway (Fig. 7.26).

One of the earliest proposals to segregate the complex arrays of visual cortical areas into “what” and “where” pathways came from Leslie Ungerleider and Mortimer Mishkin. Lesions in TE area of primates (an area homologous to IT area in humans) are known to produce serious impairment in object discrimination. One reason why neurons in the IT area are able to recognize complex objects is the large size of their receptive fields. Thanks to their expansive receptive fields, neurons in this area are also able to recognize objects with a certain translation invariance, i.e., these neurons can recognize the objects irrespective of their spatial location in the neurons’ receptive fields. There is an obvious downside to this visual strength—the neurons are unable to code for the precise spatial location of the objects. Coding for the spatial location of

objects seems to be the special function of the posterior parietal cortex. Just as lesions in the inferior temporal cortex produced impairment in object discrimination, lesions in posterior parietal cortex caused an impairment in understanding spatial relationships among objects. Monkeys with posterior parietal lesions had difficulty in choosing a response location based on its proximity to a visual landmark. Thus, the “what” (ventral) and “where” (dorsal) pathways seemed to code for object identity and object location, respectively.

The interpretation of dorsal and ventral visual pathways as the “where” and the “what” pathways was further modified by Melvyn Goodale and David Milner, who shifted the focus from what the pathways do to the inputs, to what they do to the outputs. In Ungerleider and Mishkin account, dorsal, and ventral pathways are so named based on what they do with the visual inputs—the former, the occipitoparietal system, uses visual stimulus to extract object location while the latter, the occipitotemporal system, infers object identity. In the new interpretation of Goodale and Milner, the significance of the “what” pathway remained intact. The novelty comes in interpretation of the “where” pathway, which they renamed the “how” pathway, considering its contributions to visuomotor function.

Interestingly, the new interpretation can be arrived at based on patient studies that have been earlier used to support Ungerleider and Mishkin proposal. Patients with damage to occipitotemporal region, for example, often suffer from visual agnosia, an inability to recognize common objects, faces, or abstract designs. They can, however, navigate through the visuospatial world with considerable ease. On the contrary, patients with damage to posterior parietal region suffer from *optical ataxia*, a condition characterized by the difficulty in reaching visual targets. Furthermore, they also have difficulty in adjusting the fingers, and orienting their hands with respect to the object they wish to grasp. These patients, however, have no difficulty in recognizing the visual targets.

Similar impairments were observed in patients with *Balint's syndrome*, a condition that involves bilateral damage to parietal areas. Since posterior parietal areas are responsible for understanding spatial relationships, these patients suffer from serious impairment in visuospatial skills. For example, they exhibit *simultagnosia*, an inability to perceive the visual field as a whole. They see the world as a collage of objects but cannot grasp the wholeness of the visible world. They also suffer from optical ataxia. In one study with a Balint's patient, it was observed that the patient was able to understand line diagrams of common objects. But when she tried to reach for a small object, there was no relation between the gap between her finger and thumb, and the size of the object. Therefore, damage to parietal area seems to impair a person's ability to use information size, shape, and orientation to control actions that involve acting upon that object.

## Beyond the Visual Cortex

The “what” and the “how” pathways represent, in a sense, the highest levels of bottom-up processing of streaming visual information as it ascends the visual hierarchy from the retina and flows along the network of visual cortical areas. This information is then transmitted to cortical areas in the anterior brain, where it is used for controlling action, attention, planning, and other motor and executive functions. Although we speak of a set of cortical areas as “visual” and other as “motor” or executive, in reality the labels that distinguish different areas may not be thought to define hard, inviolable boundaries. There are, for example, neurons that respond to visual information in premotor area, a cortical area in the frontal lobe that is involved in visually driven motor function. The premotor cortex receives inputs from posterior parietal cortex which in turn is activated by higher visual cortical areas, as we have seen in earlier sections. Therefore, the premotor cortex receives fairly direct inputs from the visual cortical areas. The functions of the premotor cortex and other motor cortical areas will be discussed in Chap. 9 dedicated to motor function.

## References

- Ali, M., & Klyne, A. (1985). *Vision in vertebrates*. New York: Plenum Press.
- Arditi, A. R., Anderson, P. A., & Movshon, J. A. (1981). Monocular and binocular detection of moving sinusoidal gratings. *Vision Research*, 21(3), 329–336.
- Baker, C. L., Hess, R. F., & Zihl, J. (1991). Residual motion perception in a “motion-blind” patient, assessed with limited-lifetime random dot stimuli. *Journal of Neuroscience*, 11(2), 454–461.
- Blasdel, G. G. (1992a). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, 12(8), 3139–3161.
- Blasdel, G. G. (1992b). Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *Journal of Neuroscience*, 12(8), 3115–3138.
- Conover, E. (2016). Human eye spots single photons. *Science News*. Retrieved 2016-08-02.
- Darwin, C. (1859). *On the origin of species* (p. 172) (quote on the evolution of the eye).
- Dawkins, R. (1986). *The blind watchmaker: Why the evidence of evolution reveals a universe without design* (p. 93). New York: W.W. Norton and Company.
- Ellis, H. D., & Florence, M. (1990). Bodamer’s (1947) paper on prosopagnosia. *Cognitive Neuropsychology*, 7(2), 81–105.
- Futterman, S. (1975). Metabolism and photochemistry in the retina. In R. A. Moses (Ed.), *Adler’s physiology of the eye* (6th ed., pp. 406–419). St. Louis: C.V. Mosby Company.
- Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. USA: Oxford University Press.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.
- Gross, C. G. (1973). Visual functions of inferotemporal cortex. In *Visual centers in the brain* (pp. 451–482). Berlin, Heidelberg: Springer.
- Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., ... Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, 88(5), 1621–1625.
- Hegdé, J., & Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience*, 20(5), RC61.

- Ito, M., & Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *Journal of Neuroscience*, 24(13), 3313–3324.
- Kreimer, G. (2009). The green algal eyespot apparatus: A primordial visual system and more?. *Current Genetics*, 55(1), 19–43. <https://doi.org/10.1007/s00294-008-0224-8>. PMID 19107486.
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126–1141.
- Le Vay, S., Wiesel, T. N., & Hubel, D. H. (1980). The development of ocular dominance columns in normal and visually deprived monkeys. *Journal of Comparative Neurology*, 191(1), 1–51.
- Lindberg, D. C. (1981). Alhazen and the new intromission theory of vision. *Theories of vision* (Chapter 4, pp. 58–67). The University of Chicago Press.
- Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, 194(4262), 283–287.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, 6, 414–417.
- Nilsson, D.-E., & Pelger, S. (1994). A pessimistic estimate of the time required for an eye to evolve. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 256(1345), 53–58.
- Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409(6823), 1040.
- Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecky, M. (1997). The interaction of spatial and object pathways: Evidence from Balint's syndrome. *Journal of Cognitive Neuroscience*, 9(3), 295–317.
- Sacks, O. (1996). To see or not to see. In *An anthropologist on mars* (pp. 108–152). New York: Random House.
- Savino, P. J., & Danesh-Meyer, H. V. (2012). *Color atlas and synopsis of clinical ophthalmology—Wills Eye Institute—Neuro-ophthalmology* (p. 12). Philadelphia: Lippincott Williams & Wilkins. ISBN 978-1-60913-266-8. Retrieved November 9, 2014.
- Sim, N., Cheng, M. F., Bessarab, D., Jones, C. M., & Krivitsky, L. A. (2012). Measurement of photon statistics with live photoreceptor cells. *Physical Review Letters*, 109, 113601.
- Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology*, 43(6), 482.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, 262(5134), 685–688.
- Tessier-Lavigne, M. Visual processing by the retina. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (Vol. 4, Chapter 26). New York: McGraw-Hill.
- Wong, D., & Kwen, B. H. (2005). Shedding light on the nature of science through a historical study of light, redesigning pedagogy: Research, policy, practice.
- Wurtz, R. H., & Kandel, E. R. Central visual pathways. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (Vol. 4, Chapter 27). New York: McGraw-Hill.
- Young, M. P., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, 256(5061), 1327–1331.
- Zeki, S. (1990). A century of cerebral achromatopsia. *Brain*, 113(6), 1721–1777.
- Zeki, S. (1991). Cerebral akinetopsia (visual motion blindness): A review. *Brain*, 114(2), 811–824.

# Chapter 8

## Feeling the World



*At the touch of love everyone becomes a poet.*

—Plato.

Like most other children born to well-to-do families, Padma had, or at least it so seemed, a normal childhood, but for a few crucial events that changed everything, and pushed her helplessly down a path of quiet self-destruction. Even as a 6 year old, she showed a natural flair for *Bharatanatyam*, a traditional Indian dance form. But art and music flowed in their family with many of her cousins making it big in the world of culture. Padma's father took pride in his daughter's abilities. He waited for the day when Padma would be recognized as the most talented danseuse not just in their considerably extensive family network, but beyond in the world at large. But once Padma entered high school, the growing distractions began to draw her away from the dance floor. Her performances began to wane. The promise that she originally showed now grew faint with every passing year. By the time she passed the crucial 10th grade, classical dance turned out to be the last thing on her mind.

Just around that time, Padma lost her father for an unforeseen ailment. The guilt of being somehow responsible for her father's death began gnawing away at her core. She should have been a more responsible daughter, more sensitive to her father's expectations—she heard a voice inside constantly admonishing her. Although outwardly she tried to go on with her studies as usual, inwardly she was yielding to depression, sinking under the burden of unforgiving self-condemnation. Sometime during the years of 12th grade, she set out on a path of self-imposed catharsis, though perhaps not consciously, and began to reduce her food intake. On occasions, when she was tempted to eat more than her usual, she administered laxatives in an attempt to purge her system of the extra food. Being under a state of perennial starvation, her energies began to sag. She did not have the stamina to meet the difficult academic demands of 12th grade. Her academic performance, prodigious in her early years, was now barely respectable.

With some difficulty, Padma managed admission in a less known local college. Now in her late teens, she began to grow increasingly conscious of her appearance and the consequential social acceptability. She strove to attain the legendary zero size, though she has been operating at “sub-zero” for several years now. Her emaciation shocked some of her close friends, who did not fail to notice her ribs showing in her back. Padma protested and began to use a weighing scale and a tape to make a more objective self-assessment. But these measuring devices somehow told her a different story: she was overweight and over waist. She needed to watch her diet, cut down her intake further. The objectivity that a standard measuring device is supposed to bring to one’s perceptions of the world, profoundly failed in her case. Her woes continued.

Padma suffered from a disorder called *Anorexia Nervosa*. It is a debilitating eating disorder in which the sufferer makes abnormal, voluntary attempts to lose weight, sometimes to the point of starvation. Although malnutrition is seen in anorexia, it is self-imposed and is different from malnutrition caused by poverty and lack of food. It can cause dangerous reduction in body potassium levels leading to abnormal cardiac rhythms. The most crucial and noteworthy feature of anorexia is that it is caused by a distorted perception harbored by the subject toward his/her body. This misguided perception does not seem to be amenable to correction by objective measuring instruments like scales or measuring tapes. Therefore, it is not exactly a metabolic disorder, or a dysfunction of the body physiology which exists but more as an effect. The root cause of anorexia is a brain system that maintains our body image that has gone haywire.

We do not need the constant aid of a life-sized mirror to create in our mind a constant picture of the size and shape of our body, its borders, and the manner in which it *feels*. When I stretch my hand out by activating the muscles of the shoulder and the forearm, I know and can anticipate the extent to which my hand stretches out in front of me. When I walk briskly, the returning shocks of the road on my soles make me the full weight and momentum of the body as it moves forward. When I try to stand up suddenly, I can feel the full weight of my body as it rises up from the sitting posture. There is an extensive network of sensors concealed under my skin, hidden in the recesses of my muscles, in my joints, and in my internal organs, that feed the brain with rich streaming information of the state of the body, enabling the brain create a dynamic and live image of the entity that I call myself, my corporeal self. The supportive evidence that I gather, as I gaze at myself in a mirror, descrying with a touch of disappointment an extra curve at an odd place, is completely consistent, under normal conditions, with this dynamically created internal image of the body. Trouble begins when what you see does not match with what you feel, anorexia being a rather dramatic instance of such misalignment.

What exactly is this intriguing neural system that creates and maintains a dynamic image of the body? What is its architecture? What are its different modules in the brain? How exactly do the sensors that supply it with the informational raw material work? These are some of the questions that we discuss in this chapter.

## A Philosophical Touch

In the Western tradition, one of the earliest commentaries on the nature of sense, specifically the sense of touch, can be found in Aristotle's *De Anima*. Aristotle tends to treat touch as some sort of basic sense. All animals, he says, for example, possess the sense of touch, which is presupposed by other senses. In addition to feeling touch, animals also can feel pleasure and pain, and are endowed with desire. Some animals have more specialized senses like sight, hearing, taste, and are capable of recognizing complex objects. Some animals, furthermore, have the powers of memory, imagination and voluntary movement, says Aristotle.

While Aristotle's depiction of touch in metaphysical sense, simply treats it as a basic sense, that precedes other senses like sight or hearing, his treatment of the touch sense in his other eminent ethics, brings in a certain value judgement. In this work, he describes touch as a base and inferior sense, and expresses particular contempt toward erotic touch, which he compares with "bestial" pleasures of taste. Marsilio Ficino, a fifteenth century Italian scholar who revived Neoplatonism, expresses approbation of touch on similar lines. He associated touch with more inferior and depraved forms of love, while vision was associated with more sublime, noble, and spiritual forms of love.

Touch receives a very different treatment in ancient Indian metaphysics. The difference lies in the peculiar relationship that Indian metaphysics posits between the world of matter and the senses of perception. In modern physiological and physical theory of senses, the world of matter is reduced to certain chemical and atomic constituents, whereas the sensory organs are complex and composite electro-mechanico-chemical structures that receive physical influences represented by various forms of energy, process them, and transform them into specialized sensory signals. By contrast, in Indian metaphysics, the world of matter is thought to be resolvable into five elements or elemental substances described in graphic terms as earth (*prithvi*), water (*apas*), fire (*agni*), air (*vayu*), and ether (*akash*). The idea of reducing matter to five elements permeates other ancient cultures also, like for example, the Greek and Chinese traditions. The five senses (sight, touch, hearing, smell, and taste) are thought to have a peculiar one-to-one correspondence with the five elements.

Such an association, when seen from the modern physical perspective, raises deep difficulties, since, first of all, the so-called elements are not really fundamental material principles. In response to the criticism that there is nothing elementary about earth, fire, etc. since in contemporary physical perspective, they are all highly composite principles, some have proposed that the names given to the elements (earth, water, etc.) are only metaphorical and must not be taken in their literal sense. In a commentary of the Kena Upanishad, an ancient Indian philosophical text, Sri Aurobindo presents an interpretation of the five elements that are follows. The five elements, according to this interpretation, are not really constituents of matter, but five operations, or formative stages, by which consciousness becomes matter. In the first stage, a vibration is generated in consciousness which is the basis of all creative formation. This stage corresponds to the ether or *akash*. In the second stage, there is

an immixture of the vibrations thus generated, a stage that corresponds to the “air” element or vayu. In the third stage, the vibrations organize themselves into groups, corresponding to the “fire” or agni; in the fourth stage, the “constant upwelling of the essential force” supports the form that is generated, a stage that corresponds to water or *apas*; in the last or the fifth stage, the form just generated is maintained by “an actual enforcement and compression of force”, a stage that corresponds to the earth element or *prithvi*. Thus, the elements from ether, the subtlest, to the earth, the grossest, represent the five stages by which consciousness becomes material. According to Indian metaphysics, to each of the five elements, which are objective principles, there are corresponding subjective principles, which are the five senses. Accordingly, ether corresponds to the sense of hearing (*sabda*), air to touch (*sparsa*), fire to sight (*drishti*), water to taste (*rasa*), and earth to smell (*gandha*), respectively.

Just as the theory of elements, found in many ancient cultures, remains irresolvable and isolated from the detailed depictions of the physical world in contemporary science, the theory that the elements, the bases of the objective world and the senses, the bases of the subjective world, have a unique correspondence, an intriguing idea presented by ancient Indian metaphysics, remains inexplicable and irreducible to modern science.

Thus, in Indian tradition ideas about the sense of touch or the theory of sensory perception, in general, as described in metaphysics, present nearly insurmountable difficulties to the modern thinker, resisting integration into modern science. But the sense of touch or the act of touching, as it is viewed in social and religious context, is a totally different beast. Traditionally in Hinduism, people are divided into four sections or castes, with the priestly clan considered to be the most “superior.” Social “superiority” is thought to be characterized by a certain inherent “purity,” a purity that is believed to be marred by coming into contact or touching those who reside on the lower rungs of social ladder. Therefore, in social exchanges touch carries in it an inherent peril of “polluting” the one who is touched. Over the centuries, these religious beliefs became encrusted into the perverse practice of untouchability according to which people belonging to certain castes were proscribed from coming into contact with other members of the society. Historically, other cultures of the world have also branded certain groups as untouchables—the Burakumin of Japan, Cagots in Europe, or the Al-Akhdam in Yemen. Perhaps these disconcerting religious undertones of touch in the social domain have unconsciously influenced and impeded progress in touch research in the domain of science also.

## Neglected Touch

In contrast to vision research, a scientific study of the touch sense was neglected for a long time. Taking stock of scientific literature on the touch sense in the middle of the last century, Frank lamented the poor attention paid to touch research. When a similar stocktaking was done by Matthew Hertenstein and colleagues sometimes in the last decade, the situation was found not to have improved much over the last

half a century. A systematic search of PsycINFO, a standard database of literature in psychology, revealed that the number of scientific articles published in vision in contrast to touch differed by a ratio of 13:1. A similar comparison of the articles on hearing versus those on touch produced a ratio of 3:1. What are the reasons behind this blatant neglect?

We have mentioned some reasons that are rooted in the traditional notions of purity and pollution associated with touch. Although more modern philosophical perspectives have transcended these orthodox notions, they have shown a tendency to enthrone vision vis-a-vis touch, purely from the point of view of the relative significance of the amount of information about the world that can be gathered by these two modalities. This perceived superiority of vision over other senses is echoed, for example, in the comments of Des Cartes: “All the management of our lives depends on the senses, and since that of sight is the most comprehensive and the noblest of these, there is no doubt that the inventions which serve to augment its power are among the most useful that there can be.” There are also more down-to-earth anatomical reasons that highlight the relative importance of vision compared to touch: while 30% of the cortex is dedicated to vision, and touch and hearing get only 8 and 3%, respectively. Then, there are privacy issues involved in studying touch. Touching an individual involves invading his/her personal or intimate space. There are again other technical difficulties involved in studying touch. Touch is not a single monolithic sensory modality. There are a large number of submodalities or aspects of touch that must be handled deftly if one hopes to obtain meaningful scientific results. Ultimately, the scientific community was able to overcome the religious proscription, the philosophical condescension, the social inhibition, and make rapid progress in the field of touch over the last three or four decades.

## Touch in Human Interaction

Our contacts with the physical world, the touches of the world, are mediated by our skin which envelopes and protects us like a fortress. Touch is only a broad, umbrella term that describes the kind of information conveyed by the skin. There are other types of information like tickle, temperature, vibration, and certain kinds of pain also that are carried by the receptors in the skin. Our skin is the largest organ in the body, comprising about 15–20% of the body by weight. An average human body is wrapped inside about 18 ft<sup>2</sup> of skin, innervated by around 5 million nerve endings.

Human relationships are influenced to a much greater extent than we would naively imagine. Human interactions are colored and shaped by how we touch each other. The simple physical act of touch was abstracted creatively and inserted in language to convey abstract aspects of human interactions. We speak of “rubbing” another on the wrong side. We try to keep in “touch” with those that matter, or lament, should we fail in that attempt, that we were “out of touch” with them. A superficial relationship is only “skin-deep” and an insensitive person is “thick skinned.” Something that is abhorrent or horrific gives us the “creeps,” and a fearsome encounter sends “chills

down our spine.” An agitated argument with someone could be a “heated” exchange and a cordial chat may be said to be “warm.”

The profound effect that even a casual touch can have on a total stranger has been demonstrated in an elegant study. Researchers Heslin, Rytting, and Fischer performed an experiment in a library in Purdue University. When the borrowers came to the library to present their library cards, the library clerks were instructed by the researchers to touch some borrowers and not touch others. Later the borrowers were interviewed about how they felt about the library and the staff. It turned out those who were touched, even though they were unaware or did not remember that they were touched, were the ones mostly who had positive feelings about the visit to the library and the encounters they had therein.

Humans touch each other in a rich variety of ways communicating a great depth and breadth of meaning. In one field study, Morris identified 457 different ways of touch among humans. A mother fondly cuddles and envelops her newborn as if to make the little being a part of her very self. The reassuring hug of a parent to a child who is afraid, and a light slap of the parent to the very same child as an act of admonition convey very different things. The amorous and intimate touch of a lover conveys passion. The soft touch of a master’s hand on a disciple’s head is an act of benediction or compassion. In Indian tradition, often children or younger individuals touch the feet of an elder in an attitude of reverence, while the elders bless them by placing their hand on their heads.

Let us now explore the role of touch in human interactions at various stages of our lives and in various social contexts.

## Infants Need Touch

More than half a century ago, Harry Harlow at the University of Wisconsin performed classic experiments on the effect of touch in baby monkeys on their subsequent development (Fig. 8.1). He made two surrogate mothers—one made out of terry cloth and other out of wire mesh—and placed them close to baby monkeys. For some baby monkeys, the terry surrogate provided milk, and the wire surrogate did not. With other monkeys, it was other way around. It was observed that the infant monkeys preferred the terry mother over the wired version with milk, showing that they preferred soft touch over the nutrition supplied by the milk. In most cases, the monkeys would cling to the cloth mother but lean over the wire mother to sip some milk. Most importantly, it was noted that monkeys that did not have any mother, real or surrogate, did not develop normal grooming behavior. More recently, using a similar experiment, Suomi showed that the crucial sensory stimulation that the monkeys were missing in their mother’s absence was tactile stimulation. Monkeys that were reared such that they can see their mothers behind a plexiglass, hear and smell the mother, but cannot touch her, exhibited a drastic breakdown in the development of their immune systems.



**Fig. 8.1** Harry Harlow performing experiments with baby monkeys using surrogate mothers

In human infants, it has been observed that touch can have an analgesic effect. One study by L. Gray and colleagues considered the effect of touch on infants undergoing heel lance procedure. Heel lance, which involves making a small cut in the heel, is a minimally invasive and convenient procedure for drawing capillary blood sample from a neonate or an infant. Infants were divided into two groups—one group was held by their mothers in a close whole body contact, while those in the other group were swaddled in a crib. It was noticed that infants who were held by their mothers cried 82% less, grimaced 65% less, and had a lower heart rate. Similar observations were made when heel lance was made during breastfeeding.

In a similar study Weiss, Wilson, Seed, and Paul had studied the effect of touch (harsh or soft) on 3-month-old infants, on their subsequent behavior and social adaptation as 2 year olds. They observed that infants who received harsher and more frequent touches showed more destructive and aggressive behaviors at 2 years, than those who received softer and nurturing touches. But the methodological difficulty with the study is the long intervening period between the time of stimulation (the cause) and the time of observation (the effect).

An important example of the beneficial effects of touch on an infant is massaging. Popular literature on baby care extols the positive effects of massaging on infants. Mothers are often advised to give massage to their young ones at the time of bath or otherwise, with a special baby oil. The method and mode of massaging could be a matter of convenience to the mother and comfort to the baby. But both pediatrics and popular maternal wisdom agree on the value of massaging in an infant's development. Massaging is thought to aid the infant's digestion, boost muscle development, ease teething pains, and has the effect of putting a fussy baby to sleep.

In summary, there is extensive evidence that suggests the positive and healthful influence of touch on both physical and emotional development of an infant.

## Touching Adults

Compared to the world of infants, touch has a more complex and multidimensional significance in the world of adults. The intent and meaning of touch can depend on many factors like culture, gender, and social status. Touch can be used as a powerful means of communication. The staying touch of a parent who is taking a toddler for a walk, the passionate caressing touches of a lover obviously mean very different things. However, unlike the visual and auditory channels of communication that are more open, the tactile channel is much more guarded. Communication through touch has to cross greater barriers to get across, since adults defend their personal space quite aggressively and tactile communication implies a violation of that personal space. In this regard, Stephen Thayer comments: “touch is a signal in the communication process that, above all other communication channels, most directly and intimately escalates the balance of intimacy... To let another touch us is to drop that final and most formidable barrier to intimacy.”

Touch implies arrival at the last frontier of personal space. The field of *proxemics* divides the proximal human space into several layers. The outermost of them is the public space (12–25 ft) from where public speaking is conducted; the next inner space is the social space (4–7 ft) into which acquaintances are permitted; then comes the personal space meant for friends and family (1.5–2.5 ft); and the last and the inmost layer is the intimate space (1–2 cm) wherein lie the portals of touch. Although the dimensions of these proximal spaces are determined for all people, men and women differ in their interpretation of touch.

Systematic gender differences were observed in interpreting the emotional and affective content of touch. In one study, Nguyen, Heslin, and Nguyen asked a group of subjects, mostly college-going students, to describe what it meant to them to be touched on 11 different parts of the body. Different types of touch—patting, squeezing, stroking, etc.—were considered. Touch was always delivered by an individual of opposite sex. The individuals were asked to rate the touch in response to two questions: (1) was it sexual in nature?, (2) was it warm, playful, and friendly? In case of women, the more they felt the touch was sexual, the less they found it to be warm, playful, and friendly. In case of men, it was the opposite: the more they felt the touch was sexual, the more they found it to be warm, playful, and friendly.

In a study conducted in a hospital, it was noted that 85% of the patients who were touched responded positively about the hospital and its staff, where a similar response was elicited from only 53% of the untouched patients. A more careful study was conducted by Fischer and Gallant who analyzed the gender differences in such responses. In their study, women who were touched reported lesser levels of anxiety about surgery than women who were not. Contrarily, men who were touched reported heightened anxiety about the surgery than those who were not.

Gender differences were also observed while humans interacted in public spaces. Based on her pioneering studies on gender differences in touch, Harvard researcher Nancy Henley noted that men initiated touch more often than women. There was an attempt to explain the asymmetries in touch, not merely in terms of gender inequalities, but more deeply in terms of social status. Henley proposed that touch is used as a means of communicating and establishing social status. Those in higher status initiate touch to maintain their superior position with respect to those in the lower status. Since men have an overall greater status in the society, tactile interactions show a gender bias in favor of men.

There are studies that show that gender asymmetry is mitigated by the intervention of other factors. Or, presenting it differently, even if social status is the determining factor of touch initiation, the social status of men versus women is not always fixed, but varies with the setting. Brenda Major and colleagues performed similar studies and specifically analyzed the effect of the social setting on the gender bias. Their studies have shown that while men initiated touch more often than women in public settings, in leave-taking settings (in airports or bus stations), or in recreational settings (outdoor beaches, parks, etc.) gender asymmetry was hardly present.

Brenda Major proposed that gender asymmetry favoring men is seen in interactions among strangers; but a greater symmetry will be seen in opposite-sex interactions among family members or friends. Some studies showed evidence for differential changes in gender asymmetry seen in tactile interactions. Among married or romantically involved couples, who were married or engaged for less than a year, men initiated touch more often; but among couples who were married for longer than a year, women initiated touch more often. Therefore, it appears the factors that decide the initiation of touch are more complex than mere social pecking order—it varies with gender, with the environment, and so on, a more comprehensive theory of what factor or combinations of factors determine the initiation of touch is still awaited.

Cultural factors also determine who is permitted to touch whom, factors that vary from culture to culture. Margaret Mead conducted a classic study that links the extent of touch in childhood in a culture and the levels of aggression in the adult individuals of that culture. This study was conducted in Arapesh and Mundugumor, two culturally distinct regions in Papua New Guinea. Infants in Arapesh have a lot of close contact with their mothers, who carry their young next to their skin in a small bag and breastfeeding them as and when required. The adults in their community are docile and nonaggressive, showing no cultural leanings toward warfare. By contrast, infants in Mundugumor are carried by their mothers in their baskets, held out at a distance, out of contact with their bodies. Adults in this culture are more aggressive, prone to warfare and violence. Likewise, it was noted that the Kung babies of the Kalahari in Africa, who are carried by their mothers in a close skin-to-skin contact, also grow up to be peaceful adults.

The level of contact in the form of greetings also varies among world cultures. A wide variety of acts and gestures including handshaking, nose pulling, hair tousling, kissing, cheek tweaking, head patting, back slapping, kissing, and embracing can be involved in a greeting. Greetings in certain cultures involve closer contact while

the prevailing etiquette in others prescribe a safe distance, and the contact may not go beyond a light handshake. Traditional greeting in India and surrounding regions involves a respectful gesture of *Namaste* with folded hands and no body contact. Australian friends are known to kiss and even cry over one another. Greek and Italian greetings involve considerable touching where strangers to a home are greeted with a strong embrace and a kiss on the cheek. In the Arab world, greeting in a formal setting involves a handshake; however, among family members and friends, it usually involves an embrace and a kiss on either cheek. Generally, societies living in the Mediterranean nations (Spain, Italy, Greece, France, etc.) use high contact greetings while the societies of the northern Europe (like Holland or UK) and North America (USA) prefer to touch less.

The isolationist tendency seems to be particularly strong in the British society and a lot has been written about it. Perhaps the famous lines (“No man is an island, Entire of itself, Every man is a piece of the continent, A part of the main”) from a poem by John Donne, an English metaphysical poet, is a pointer to this aspect of English social attitudes. Ashley Montagu comments on how touch is discouraged and repressed in English culture: “England is a land full of peculiar people, of people who are adults, who seldom touch each other, and in which one apologizes to one’s father or one’s mother when one touches them accidentally. This, of course was a rule in well-bred families which means more care in breeding horses than care in breeding children.”

So far, we have seen aspects of touch in social interactions, we have considered the many social colors of touch. We have seen the nurturing role of touch in an infant’s development. We have seen that in social contexts, the differences in status determine the asymmetries involved in tactile interactions. Even gender-related asymmetries in tactile interactions have been linked to social status. One aspect of touch that influences the gender asymmetries, particularly, in cross-gender interactions is the aspect of sexuality. Certain touches are interpreted as sexual while certain others are not—a difference that unfortunately does not lend itself to easy objective discrimination. In fact, it is this aspect of touch, as opposed other sensory modalities like vision or audition, that had historically placed barriers in progress of touch research. We have also noted the considerable cultural differences in touch. Cultures that encourage touch, particularly, in early development, seem to be more at peace with themselves and with others. Higher levels of aggressiveness were observed in cultures in which there was a general, subtle proscription of touch.

In the following section, we will consider the basic machinery of touch. How are the physical contacts of the world that we interpret as touch transduced by the body? What are the different categories of touch and how are they processed by the body’s tactile sensory apparatus? By what pathways and stages are these signals transmitted to the highest touch regions of the brain? How are the tactile signals streaming from various parts of the body integrated into a whole called the body image? These are some of the questions that will be addressed in the following section.

## The Engines of Touch

In the summer of 2006, Kojiro Hirose organized a unique exhibition as a part of the National Museum of Ethnology in Osaka. The museum seeks to spread awareness about different societies and cultures of the world through collection and curation of ethnographic materials. The theme of the exhibition was—"Touch and grow rich: You can touch our Museum." It was created in the spirit of Kuzuhara Koto who created the first Japanese Braille in the nineteenth century. Unlike the special dot patterns used in the Western Braille, Kuzuhara Koto's Braille printed natural Japanese characters in an embossed form, the intention being that both the blind and the sighted would see or read similar text (Fig. 8.2). In a typical museum, the visitors would merely see the exhibits and are actually often forbidden from touching the exhibits. The distinctive feature of the exhibition organized by Kojiro Hirose was that the exhibits were meant to be touched and felt and not so much seen. It was visited by large numbers of visually handicapped people who traveled from far off places in Japan to see the exhibition. The exhibits included historical writing systems for the blind like raised wooden letters, needle letters, origami paper letters, and so on. There was a famous Fureai Buddhist statue, delicately carved bird sculptures, and models of entire Japanese shrines. The visitors touched exhibits, passed their hands over them, held them in their hands, turned them over and over, and experienced every tactile detail of those objects.

When you explore the world through touch, when you process the tactile information streaming in, for example, from a tiny Buddha figurine that you are holding and turning over in your hands, to identify the object, appreciate the sculptural aesthetics, you are using the full power of your somatosensory system, the part of the



**Fig. 8.2** A sample of Japanese "Braille" or raised letters (*totsu monji*) from the collection of Kyoto Prefectural School for the blind

brain that process the touch and a host of other forms of allied sensory information, denoted by the umbrella term known as the somatic sense. Touch is what happens when an object presses against your skin, an event that produces a response in your brain. Here again literature distinguishes between passive and active touch, though the distinction is rather gray. When you receive the contact of an object passively, or press your finger, say, on an embossed letter and identify it, it is passive touch. But when you move your fingers over a surface, and actively explore it to understand and interpret it, you are engaged in active touch.

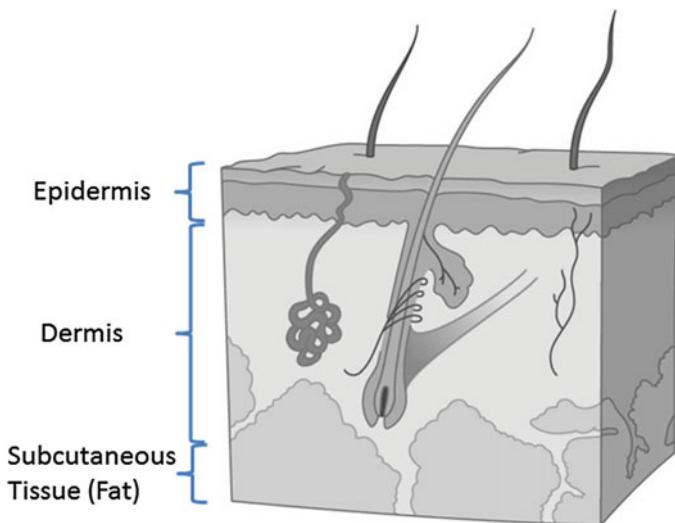
Then, there are distinct sensors that detect different aspects of touch. There are sensors that specifically respond when you press softly on an object with your fingers; there are sensors that respond when you gently pass your fingers over an object; there are sensors that detect the vibration of a surface; there are sensors that permit your brain to recognize the object you are holding, based on the configuration of the fingers of your hand; and there are sensors for feeling the warmth of the object you are holding.

When an object comes into contact with your skin, it leaves a temporary, object-shaped imprint in your skin. This indentation of your skin activates certain receptors embedded in the skin. There are four types of sensors or *receptors* in the skin. These are called *mechanoreceptors* since they transduce mechanical contacts with the external world. The four mechanoreceptors are Meissner corpuscles, Merkel cells, Pacinian corpuscles, and Ruffini endings. These are tiny specialized structures that can transform a mechanical impact into an electrical signal. The four mechanoreceptors are distinguished by the depth at which they are present under the skin and the type of axons that innervate them. Mechanoreceptors located deeper in the skin are harder to stimulate, but can integrate tactile stimuli acting over a larger skin surface. The axons that innervate the mechanoreceptors too are classified as rapid adapting or slow adapting based on their speed of response to stimuli. Thus, there are two dimensions that distinguish the four mechanoreceptors—the area of influence and the speed of response.

The mechanism by which the mechanoreceptors transform mechanical stimuli to electrical signals can be explained in terms of the special ion channels they possess. The deformations that mechanoreceptors undergo, on application of pressure, stretches certain special stretch-sensitive ion channels, which open and allow passage of current, thereby converting a mechanical signal into an electrical signal.

The skin is made of three layers (Fig. 8.3). The outer layer is the epidermis whose thickness varies from 0.05 mm in the eyelids to 1.5 mm in the thicker skin of the palms. Below the epidermis there is the dermis varying in thickness between 1.5 and 4 mm, making up nearly 90% of the thickness of the skin, accommodating blood vessels, lymphatics, hair follicles, and so on. Beneath there is the subcutis, the innermost layer of the skin.

Merkel cells are small epithelial cells that wrap around axons of a sensory nerve fiber. They are found at higher density in hairless or glabrous parts of the skin, like, for example, the palm and the fingertips. These skin regions have narrow ridges visible to the naked eye. The ridges afford the skin of the fingertips a special texture. The subtle corrugations or ridges on the surface of the fingertips enable in complex

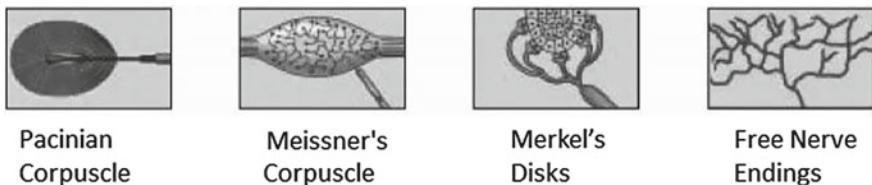


**Fig. 8.3** The layers of skin

microscopic, mechanical interactions between the fingertips and any surface that may be felt by them. The ridges increase the friction of the fingertips and the palm, making it possible to grasp objects firmly without slippage. This principle is used in the design of a lot of real-world surfaces. Bottle caps have ridges on them so as to increase friction and making it easy to open the caps. Bathroom tiles have coarse surfaces to increase friction, and thereby prevent slippage and fall. Merkel cells are found at particularly high densities under these ridges of hairless skin. Thanks to their strategic location under the ridges, Merkel cells respond best when the fingers pass over sharp edges or pointed ends. They play a role in reading Braille. They are innervated by slowly adapting axons.

Meissner's corpuscles are globular structures located in the dermis, embedded in the dermal papillae, folds of the dermal–epidermal border (Fig. 8.4). They ensheathe a set of flattened, lamellar (“planar”) cells that are mechanically coupled to the skin surface by collagen fibers. This coupling endows the Meissner's corpuscles with a high tactile sensitivity, particularly, to tangential stimuli that move across the skin surface. They are located in many areas of the skin but at higher densities in the glabrous skin of the fingertips and lips. They are innervated by rapidly adapting axons.

The Pacinian corpuscles are also located in the glabrous skin but with particular sensitivity to vibratory stimuli, capable of sensing frequencies as high as 250 Hz (Fig. 8.4). They are roughly oval shaped and are 1 mm in length. They are fewer in number than the other three types of mechanoreceptors. They are also located deepest inside the skin compared to the other mechanoreceptors, and therefore have a large receptive field.



**Fig. 8.4** The mechanoreceptors

**Table 8.1** The four mechanoreceptors and their distinctive properties

|                  | Small receptive field | Large receptive field |
|------------------|-----------------------|-----------------------|
| Slowly adapting  | Merkel cells          | Ruffini nerve endings |
| Rapidly adapting | Meissner corpuscles   | Pacinian corpuscle    |

Ruffini's nerve endings are enlarged dendritic endings with elongated capsules (Fig. 8.4). They are located inside the dermis. They are innervated by slow adapting axons and respond better to stretch of the skin than to indentation (Fig. 8.4).

Table 8.1 shows how the four mechanoreceptors occupy, respectively, the four quadrants offered by the two key dimensions of touch transduction—receptive field size and speed of response.

In addition to the above mechanoreceptors that are present inside the skin, and therefore cutaneous receptors, there are other receptors that sense properties of the skeletal muscle, and therefore contribute to the broad somatic sense. As described in Chap. 9 on motor function, these are the muscle spindles that sense the muscle length and velocity, and the Golgi tendon organs that sense the muscle tension and communicate these variables to the brain. By integrating the information about muscle lengths, the brain is able to assess the configuration of various joints. By analyzing the pattern of tensions, the brain judges the weights and masses of the objects that are interacting with the body. The sense that enables the brain to judge the joint configuration of the body is called *proprioception* or the position sense.

Then, there are nociceptors, receptors that sense pain (nocere is Latin to injure or cause pain). Unlike the mechanoreceptors that have specialized processes for transducing mechanical impact, the nociceptors are simply free nerve endings. They are widely distributed in the body in the skin, muscle, joints, and the internal organs. Since pain can be produced by a variety of stimuli, nociceptors can be activated by a variety of stimuli including mechanical, chemical, thermal, or electrical, dubbed collectively as noxious stimuli. Two kinds of axons are associated with nociceptors—the fast A $\delta$  fibers and the slower C fibers. Activation of A $\delta$  fibers produces a sharp, local pricking pain, whereas activation of the C fibers produces a more dull and diffuse pain. We know that all the differences in the nature of the energy involved in sensory stimuli (optical, mechanical, electrical, and chemical) disappear once the stimuli are transduced and transformed into a common type of signals viz., the electrical neural impulses. Therefore, even electrical stimulation of C fibers produces intense burning sensation even though no thermal stimulus is applied peripherally.

Another important class of receptors that are generally studied along with other somatic receptors are those that sense temperature—the thermoreceptors. Thermoreceptors are different from a thermometer, where the same device can sense temperatures that correspond to what we feel to be “cold” and what we feel to be “hot”. There are separate thermoreceptors for “hot” and “cold.” Furthermore, among the “hot” thermoreceptors, there are warm receptors and heat nociceptors. The range of temperatures between 31 and 36 °C is sensed to be warm, while the range from 36 to 45 °C begins to feel hot. Beyond 45 °C, the heat nociceptors begin to be activated and such heat is felt to be painful. Up to 45 °C, the warm receptors act nearly like thermometers, with their firing rate increasing linearly with temperature. Similarly, on the cooler side, as the temperature drops from 31 to 15 °C, the thermal sensation ranges from cool to cold, bordering on the painfully cold at around 10 °C.

Thermoreceptors are not confined only to skin but are found in other parts of the body also. Temperature sense in the tongue, for example, plays a key role in the sense of taste. It is a common experience that food that is hot, or at the right temperature tastes the best. Even ice cream tastes better when it is not too cold, but on the verge of melting. These points of general experience have been studied formally for nearly a century. Our threshold to sense the four basic tastes—sweet, salt, sour, and bitter<sup>1</sup>—shows a minimum around 20 and 30 °C. As the temperature exceeds 30 °C, it becomes increasingly harder to detect weak tastes. Similarly, the “burning” sensation caused by a spicy substance like the chili pepper is due in part to the warm receptors and the hot nociceptors on the tongue.

We have quickly reviewed the key somatic sensors including the mechanoreceptors and other allied receptors like the nociceptors and thermoreceptors. These are transducers that transform various forms of energy that impinge on the body surface into electrical signals that are conveyed to the brain. These signals now climb upward via the spinal cord to the cortex and other higher brain areas where complex somatosensory representations of the world are constructed. The architecture of this somatosensory system is considered below.

## The Somatosensory System

It helps to draw analogies between the visual system and the somatosensory system to understand the hierarchical architecture of the latter. In the visual hierarchy, starting from the retina to the highest visual cortical areas terminating in the “where” and the “what” pathways, we have noted a progressive development of visual representations, with simple geometric primitives (dots, lines, etc.) represented in the lower stages and more complex visual objects represented at higher levels. An architectural feature that facilitates such progressive elaboration of visual representation is the fact that each neural stage in the hierarchy “looks at” only a small window of neurons,

---

<sup>1</sup> Japanese researchers maintain that there is a fifth basic taste called umami that roughly corresponds to the taste of Monosodium Glutamate (MSG).

which constitutes its “receptive field,” in the previous stage. By virtue of such an arrangement, neurons in the higher stages have progressively larger receptive fields and are capable of representing more complex visual patterns. A similar pattern of hierarchical representation is seen in the somatosensory system also.

The representations that are generated at various stages in the somatosensory hierarchy can be best described in terms of two concepts—receptive fields and tuned responses. Again returning to the analogy with the visual system, the notion of the visual receptive field is usually applied to the neurons and not the receptors themselves. Accordingly, even in the retina, one can speak of the receptive fields of the bipolar cells or the ganglion cells, but not of the rods and cones, although it is in principle possible to talk of the region in the visual field from which a photoreceptor can receive light. But in case of the somatosensory system, the notion of receptive field is applied right from the level of the mechanoreceptors. This is perhaps because, there is a dermal layer that separates the mechanoreceptors from the external stimulus and, depending on the depth of the mechanoreceptor, it can respond to stimuli applied over an area of the skin that is more or less wide. In the hand, for example, the receptive fields are smaller in the fingertips due to higher densities of the mechanoreceptors there, than in the palm region. The superficially located, slowly adapting Merkel cells have smaller receptive fields, and respond to small dot-like pressure pattern on the skin. The Meissner corpuscles, also superficial, but rapidly adapting, have small receptive fields, but have a pressure sensitivity that is highly inhomogeneous over the receptive field. Also note that the receptive field is a static concept and does not capture the response pattern of a neuron through time. Therefore, a characterization of the response of a rapidly adapting mechanoreceptor exclusively in terms of its receptive field is necessarily incomplete. The Pacinian corpuscles, located in deeper layers, have large receptive fields, extending sometimes over an entire finger, or a large section of the palm. Furthermore, there is a single point of high sensitivity right over the Pacinian corpuscle, with surround regions producing weaker responses. Finally, the slowly adapting Ruffini’s nerve endings, also deeply situated, have large receptive fields, with a distinctive feature that their response depends, not only on the point of stimulation in the hand but also on the direction of stimulation. The resemblance to the direction-sensitive neurons in the lower stages of the visual system is unmistakable.

## Dermatomes

Drawing analogies with the visual system once again, we note that as we ascend higher in the somatosensory hierarchy, the receptive fields grow larger and more complex. Typically, the notion of the receptive field is applied to single neurons and their responses. But above we have extended the notion to single mechanoreceptors and described their receptive fields located within the overlying skin. We now describe the receptive fields that obtain at the next stage of the somatosensory hier-

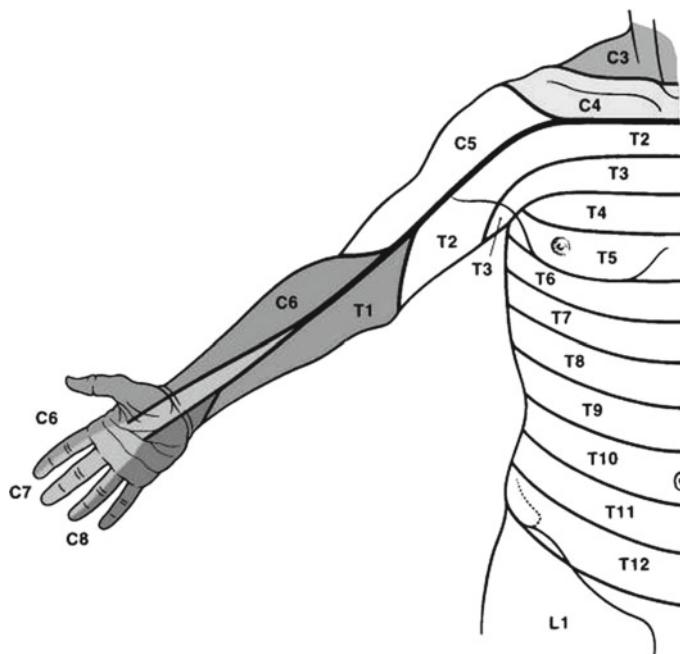
archy. Again, these receptive fields do not pertain to single neurons but to bundles of nerve fibers.

We may recall from Chap. 2 that there are 31 pairs of spinal nerves connecting the body and the spinal cord and 12 cranial nerves linking the brain and the body. The fibers of the spinal nerves are axons of the neurons located in the Dorsal Root Ganglia (DRG), located outside the spinal cord in the intervertebral foramen. The neurons of DRG are rather peculiar and deviate significantly from the textbook neuron described in Chap. 3. In a classical neuron, inputs are received at one end by the dendritic tree, integrated in the soma, and conveyed outward via the axon and its collaterals. The dendrites, typically short in span, have the machinery for receiving information—the receptors—and the typically longer axons have the machinery for sending out information, the machinery for neurotransmission. But a sensory neuron is faced with a peculiar challenge. It must have long fibers that extend from the spinal cord all the way to the peripheries. But these fibers must have receptors since they must carry sensory information. Therefore, to serve this purpose, the sensory neurons of the DRG have actually a single axon that is split into two branches. One of these branches extends out to the peripheries and terminates, at the collaterals, in the mechanoreceptors, or simply branches out into free nerve endings. The other branch passes through the dorsal root and sends out projections to various parts of the cord. Such neurons are called the pseudounipolar neurons, to distinguish them from the “pure” unipolar neurons that have only a cell body and the axonal arbor.

Nerves exiting the four main areas of the spinal cord—cervical, thoracic, lumbar, and sacral—innervate well-defined areas on the skin surface and also internal/visceral organs. The cervical nerves innervate the arm and the hand areas; the thoracic nerves innervate the head, neck, thorax, and the abdomen; the lumbar nerves innervate the lower abdomen, hips, and anterior side of the legs, leaving the foot area; and the sacral nerves innervate the posterior side of the legs and the foot area.

The region of the skin mainly innervated by the afferent (sensory) fibers of a spinal nerve is called its *dermatome* (Fig. 8.5). The dermatomes are not mutually exclusive: neighboring dermatomes overlap. A given skin area is typically innervated by three adjacent spinal nerves. Therefore, in order to anesthetize a given skin region, it is necessary to block signals of three adjacent nerves. Most spinal nerves (with the exception of C1, the first cervical nerve) are mixed nerves. That is, they have a mix of afferent (sensory) and efferent (motor) fibers. Cranial nerves carry sensory information from the special senses (eyes, nose, etc.) and motor information to the muscles of the head, neck, and muscles that control eye movements, muscles of mastication and speech articulation. Some cranial nerves carry cutaneous and proprioceptive information, and therefore rightly serve the functions of the somatosensory system. The trigeminal nerve, for example, carries cutaneous information from the skin on the anterior half of the head.

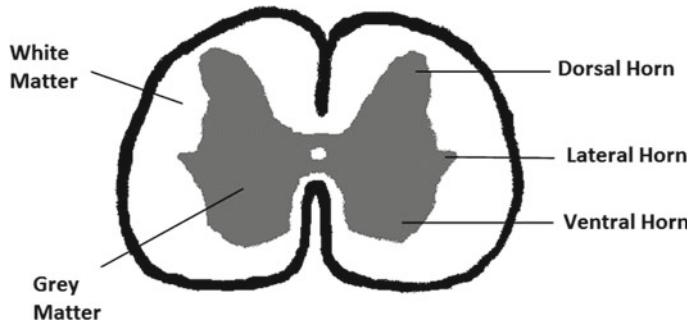
The fact that both the skin and various visceral organs are innervated by the same set of spinal segments gives rise to the intriguing phenomenon of “referred pain.” If the fibers from a certain visceral organ and those from a region of the skin project to the same set of spinal neurons, pain in the visceral organ is felt to be originating, not from the viscera itself, but from the corresponding area of the skin on the body surface.



**Fig. 8.5** The dermatomes

Thus, when brain localizes pain originating in the body, it often projects the internal sources of pain to the surface of the body. For example, people who experience pain due to myocardial infarction report pain originating from the left arm and chest. This is because, nociceptive fibers from the heart and sensory fibers from the left arm project to the same spinal segments and to the same neurons. Similarly, pain in the liver is felt in the upper right area of the abdomen just below the ribs. Pain in the spleen is felt near the shoulder at an area just above the collarbone. This clinically observed feature known as Kehr's sign was named after its discoverer, the German surgeon, Hans Kehr.

A section of the cord, anywhere from the cervical to the lumbar regions, shows a central gray matter and a surrounding white matter. Note that this is reverse to what we see in the cerebrum—cortical gray matter is all around, with the white matter projecting inward from the cortex. The central gray matter of the section of the cord is shaped like a butterfly, and the tips of the “wings” of the “butterfly” are called horns. There are totally four horns as shown in Fig. 8.6. There are two on either side of the midline; there are two each on the dorsal and the ventral side, known as the dorsal and the ventral horns, respectively. The gray matter extending from the dorsal and the ventral horns is divided into ten zones called the laminae numbered from I to X. The axons of the DRG neurons project to various parts of the gray matter of the spinal cord. Neurons of lamina I receive information about thermal, noxious, and visceral stimuli.



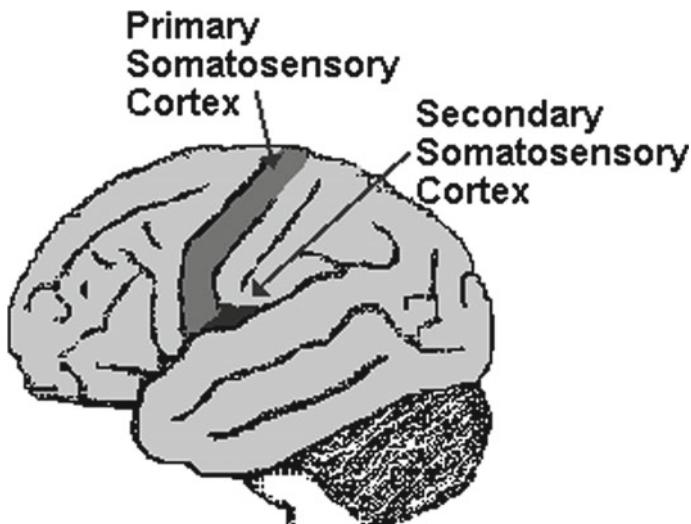
**Fig. 8.6** A section of the spinal cord showing the horns

Neurons in lamina V receive mechanical, noxious, and visceral stimuli. Neurons in one lamina may project to neurons of other lamina either at the same level, or to other spinal levels, or to the cortex via the thalamus. Signals from the spinal lamina are sent to the thalamus via two ascending pathways: the dorsal column-medial lemniscal system carries tactile and proprioceptive information, whereas the spinothalamic tract carries pain and thermal information.

## The Somatosensory Cortex

The primary somatosensory cortex (S-I) located in the postcentral gyrus of the parietal lobe is the port of entry for the somatic information that is conveyed by the spinal cord via the thalamus. S-I further comprises several subregions identified in terms of Brodmann areas: 3a, 3b, 1, and 2. Regions of S-I project to the next somatic cortical area in the hierarchy, the secondary somatosensory area (S-II) (Fig. 8.7). Areas 3b and 1 receive tactile information from the skin, whereas areas 3a and 2 receive proprioceptive information from muscles and joints. As we have pointed out earlier, as we go up the hierarchy, neurons have progressively larger receptive fields. Therefore, neurons of areas 3a, 3b, and other S-I subareas have larger receptive fields than those of the fibers that carry tactile information from the skin to the spinal cord. The S-I neural receptive fields could cover an entire fingertip, sometimes even extending over a few adjacent fingertips. The neurons of S-II, on the other hand, have larger receptive fields extending over the fingertips and the palm of either one hand, or sometimes including the corresponding regions of both hands. Perhaps this is because when we hold a large object with both the hands, corresponding areas of the two hands are simultaneously in contact with the object, and therefore logically eligible to be considered as parts of the same receptive field.

One more point of similarity between visual cortical neurons and the somatosensory neurons is that receptive fields of both neurons have center-surround structure. Stimulation in the central parts of the receptive fields of somatosensory neurons



**Fig. 8.7** The primary and secondary somatosensory cortices

results in increased firing, while stimulation in the peripheral regions causes inhibition and reduced firing. The somatosensory cortex too has a columnar structure. These columns are each roughly of size  $300\text{--}600\text{ }\mu\text{m}$ , extending all through the six cortical layers. Even though neurons in a given column may not have identical receptive fields, they share the central regions of their receptive fields.

Like in the primary visual cortex, columns of S-I also have an elegant topographic organization, with response properties varying continuously along the cortical surface. But the analogy ends there because the features that are mapped in the visual cortex are properties of visual stimulus like orientation, direction of motion, color, etc. In the somatosensory case, it is simply the location on the skin that is mapped on the cortical surface. Thus, there is an entire map of the body surface, a somatotopic map, on the cortical surface. In lay parlance, these maps are “upside down” with the legs on top and hand and face regions at the bottom. In more formal spinal terms, in the somatosensory cortex, the sacral segments are mapped medially, lumbar and thoracic segments centrally, and cervical segments laterally.

The somatosensory maps are not drawn to scale and are characterized by significant local distortions. Some body regions are allotted a disproportionately large cortical real estate. For example, the hand, foot, and mouth areas have correspondingly large cortical maps. The trunk, on the contrary, which has a large real area, has a relatively compact cortical map. The key factor that determines the amount of cortical allocation to given body region is the density of tactile receptors in the skin. The hand, foot, and mouth regions, though small in real terms, have a larger density of tactile receptors. The trunk, despite its large surface area, has relatively fewer receptors.

Within S-I itself, there is a separate whole body map in each of the subareas of S-I viz, 3a, 3b, 1, and 2. As mentioned above the four subareas of S-I are distinguished by the kind of inputs they receive from the receptors in the peripheries. Area 3a receives input from both slowly adapting and rapidly fibers, whereas area 1 receives inputs mainly from rapidly adapting fibers. Therefore, areas 3b and 1 primarily encode tactile texture like, for example, the difference between the touch of a silk fabric and sandpaper. Area 3a on the other hand primarily responds to the proprioceptive information from the stretch receptors of the muscle while area 2 responds to both tactile and proprioceptive information. Therefore, areas 3a and 2 convey information about the joint configuration. These are the brain areas that help you hold that cricket ball precisely between your fingers and the palm of your hand deliver that perfect spin. The ability to process joint configuration enables to produce such a configuration with our hand and fingers so as to hold or grasp an object. It is this ability that enables us to recognize a three-dimensional object based on how it *feels*, by turning it over in our hand with our fingers, just as visually we recognize an object by how it looks. The problem of recognizing an object from how it feels is known as haptic object recognition, a notion that is discussed further on in this chapter.

## Recognizing Objects Through Touch

When we speak of objects, we often intuitively mean the visual appearance of the object, identified by its color, form, size, and other visual attributes. But you can also identify an object purely by the way it feels, by its touch. You know quite well, even with your eyes closed, how your pet terrier feels on your lap as you cuddle its warm, soft, and furry form. You know the exact shape of the hollow of your hand as you grasp a computer mouse. Just as you identify an object by the combination of the attributes that it represents, you can also identify an object *haptically* (the technical term for “by touch”) by its surface textural properties (is it smooth, rough, sandy, corrugated, knobby, furry, etc.?), or its size (its length, breadth, and height), or its geometry (cuboidal, spherical). Textural properties are encoded in a rudimentary fashion the responses of the mechanoreceptors and processed by the neurons of S-I. Aspects of size are encoded by the configuration of the joints of your fingers that your hand assumes in order to hold the object. In other words, size is encoded in the proprioceptive information. How do we combine all these different haptic attributes to construct the image of a haptic object?

In order to answer that question, let us hark back once again to how objects are represented in the visual system. Although there are profound differences in vision and touch, neurobiologists of haptics have found it quite convenient to allow the knowledge of visual science to guide the early efforts to make sense of the underlying hierarchies of haptic object recognition. Our understanding of visual object recognition is a much celebrated success story not only in visual neuroscience but perhaps also in all neuroscience. Visual science has unraveled an elegant hierarchy of representations in the visual system from the primitive dot-like patterns in the lowest

levels, with complex objects represented at the highest levels in the inferotemporal cortex. As one ascends the hierarchy, the various attributes of an object are progressively combined, to construct a final representation of an object. A similar hierarchy has also been worked out and identified even in the somatosensory system in order to explain the underlying processes of haptic object recognition.

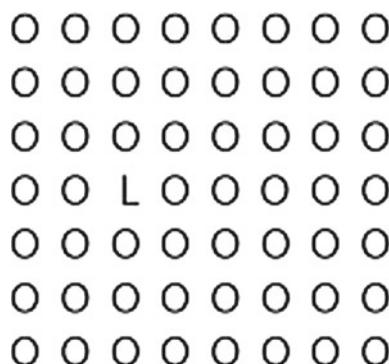
The crucial question underlying object recognition is how the brain integrates the various features that characterize the object in order to create an identity of the object? This question was answered by the Feature Integration Theory proposed by Ann Treisman and Garry Gelade in the ‘80s. According to the Feature Integration Theory, when an observer looks an object with certain features, the features are processed separately and preattentively (even before the conscious attention of the observer steps in) in separate feature maps in the brain. Since this happens preattentively, the observer does not become aware of this process of separation of object features. And, since it happens in parallel across the feature maps it happens quickly. The feature maps in turn project to a master map where specific combinations of features are detected. Next a “spotlight” of conscious attention is directed toward these combinations or conjunctions of features; feature conjunctions that are attended to are stored temporarily in “object files.” A comparison is then made between the prior knowledge of the object and the object file just generated; when there is a match, the object is identified.

The Feature Integration Theory was able to suggest experiments and make predictions in the area of visual object recognition, or rather, feature integration that forms the basis of more complex object recognition. In a typical experimental paradigm to which the Feature Integration Theory is applied, the subjects are required to determine the presence or absence of a target among an array of distractors displayed over a two-dimensional area and time taken by the subjects to determine the answer is measured. Experiments in which the target and the distractors are distinguished by a single property, the search in time is nearly the same irrespective of the number of distractors. An example of such an experiment is to search for the “L” hidden among “O”’s in Fig. 8.8. No matter how many “O”’s surround “L”, it takes about the same time to locate. The target “L” seems to spontaneously “pop” out of the image. This is because, in this case, the information that is necessary is readily available in a feature map that is constructed preattentively and quickly.

But the situation is different when there are multiple properties that differentiate the target from the distractor. Consider the problem of searching for a green “T” from an array of blue “T”’s and green “K”’s. Now, there are two feature dimensions involved: the “T-K” axis defined by the letter, and the “green-blue” axis defined by the color. In this case, the search time was found to increase linearly with the number of distractors. The result suggests that in this case a serial search process is involved in which the attentional spotlight hops from one item to the other in the array of distractors (and the target) looking for the right conjunction of features.

Similar results have been obtained even in the domain of haptic object recognition involving single and multiple features. As in the visual case, in somatic feature search also a pop-out phenomenon was observed when only a single somatic property was involved. In one such a study conducted by Myrthe Plaisier and colleagues, the

**Fig. 8.8** A single target (L) hidden among a set of identical distractors (O)



subjects were asked to explore with their fingers a flat surface on which small patches of sandpaper of varying roughness were pasted. One of the patches was a target patch which differed in its roughness compared to the remaining distractor patches. The target was present in some cases and absent in others. The subjects were required to explore the array of patches in the shortest possible time and detect the presence of the target. When the target was distinctively rough while the distractors were smooth, the search time was nearly flat as a function of the distractors. So the rough target showed a pop-out effect. However, unlike the visual case, when the target was smooth while the distractors were rough, the pop out did not occur. The search time increased with the number of distractors indicated an underlying serial search process. Thus, rough targets against smooth distractors are not searched in the same way as smooth targets against rough distractors, a feature known as search asymmetry. The differences in search times also corresponded to the patterns of exploration. Search times were shorter when subjects could detect the target in a single sweep of the hand over the array, whereas in cases where the subjects adopted more complex exploratory strategies search times were longer. A similar pop-out effect was found while searching for 3D objects like, for example, searching for tetrahedrons among spheres. Interestingly, pop out occurred even for thermal properties. In one study, the subjects were asked to grasp a variable number of spheres at different temperatures. The distractor spheres were at 38 °C warmer than the skin, while the target sphere was colder (22 °C). In this case too, the pop-out effect was observed.

The above examples of haptic search involve items defined by a single feature. It is interesting that even in case of single feature dimension, in haptic search, there is a linear increase in search time under certain search conditions. The situation is certainly more complicated in case of haptic conjunctive search involving multiple features. Some of the earliest experiments of this kind were performed by Russian (then Soviet) scientists Lomov and Vekker with the explicit aim of developing reading materials for the blind. The creator of such materials must be able to design items, ideally near-planar elements stuck to a surface that can be easily classified into a large number of haptic classes by a haptic observer. Furthermore, in order to increase the number of such classes, it might be essential to use haptic items that present

multiple features. These original motivations have inspired several conjunctive search experiments in haptic domain.

The haptic conjunctive search experiments have shown that several haptic properties are processed together or “co-processed.” For example, a study by Corsini and Pick has shown that texture and length are processed together. Textures are often denoted by “grit values” a quantity that is inversely related to the sizes of the particles whose distribution on the surface of interest defines the texture. A sandpaper of grit 20 has coarse grit and one with grit value of 1500 has fine grit. The study of Corsini and Pick had chosen items with five grit values ranging from 24 to 320 and five lengths ranging from 7 to 7.75 in. The subjects were presented pairs of items and were asked to judge the longer item. The study showed that fine texture items were perceived to be longer than the ones with a coarse texture. There is a long line of studies in the haptic object recognition that considered the interaction among multiple tactile and haptic properties like shape, texture, hardness, vibration, etc. on the efficiency of object recognition.

Above, we painted a picture of a haptic object as a passive collection of haptic properties: the subject, by coming into a passive tactile contact with those properties, readily recognizes the object. But this picture is not quite true to the reality. A real-world haptic user does not simply touch an object to understand it; he/she engages in complex exploratory interactions with the object like, for example, turning it over and over his/her fingers, in order to attain a satisfactory haptic understanding. To involve the visual analogy again, the aforementioned account of haptic object recognition is like viewing a small photograph of a person to recognize the person. A simple presentation of the whole image is probably sufficient. But we are viewing a complex scene with many visual elements, we do not process it like a snapshot. We scan the image with our eyes, focussing on various areas of interest, mopping up significant local information, before shifting attention to new areas. Such visual exploration is usually driven by fast, darting movements of the eyes called saccades. Similar movements are made by the hands as they hold and manipulate a haptic object.

Roberta Klatsky, Susan Lederman, and colleagues have studied the nature of these exploratory haptic movements extensively and have labeled them Exploratory Procedures. The researchers have perceived that certain stereotypical hand gestures repeatedly surface in haptic exploration. Although a considerable number of exploratory movements have been defined and identified, there are five that figure quite prominently and studied widely. Each of these movements is intended to extract specific haptic properties of the object being explored. The first of these movements is *lateral motion*, the kind of motion in which you stroke the surface of an object using gentle lateral movements of your fingers. Lateral motion is used to extract texture information from the object. The next movement is *pressure*, where you press on a surface with your fingers in the normal direction. The aim of this movement is to understand the compliance of the object’s material. Then comes the *static contact* where you gently press the object with your open palm so as to obtain a close contact of the object surface with your skin. Through such a movement, your skin can attain thermal equilibrium with the object, and be able to measure the object’s temperature.

Then, there is *enclosure*, wherein you grasp the object completely with the fingers of your hand. Such a movement discovers the overall shape of the object and assesses its volume. Finally, there is *contour following*, in which typically you hold the object in a fixed position with one hand, and explore the object with the fingers of your other hand. Contour following yields a more detailed understanding of the object's shape. The above five exploratory procedures are some of the primitives of haptic exploration. Actual haptic exploration of real objects involves far more complex movement patterns, the diversity of which can only be rivaled by the diversity of real-world objects themselves.

Considering the crucial role of exploratory movements, it is evident that the brain regions underlying haptic object recognition will not be confined to somatosensory areas but will necessarily involve motor cortical areas also. These expectations have been confirmed by functional imaging studies. In one such study, Catherine Reed, Shy Shoham, and Eric Halgren have asked subjects to manipulate real world and "nonsense" objects. The brains of the subjects were scanned by functional magnetic resonance imaging devices as they haptically explore the experimental objects. The real-world objects included familiar objects like Q-tip, whistle, tennis ball, apple, book, carrot, and football. The nonsense objects were unnatural 3D shapes carved out of balsa wood. The brain activation produced by exploration of real-world objects was compared to that of nonsense objects. In case of real-world objects, most prominent activation was observed in the secondary somatosensory areas of parietal operculum and insular cortex. This is expected since complex haptic objects are likely to be represented in higher order somatosensory cortices by combining more preliminary information from the primary somatosensory cortex. Lateral areas of the visual cortex, which are normally associated with visual object recognition, are also activated. It is interesting that visual areas are activated even though the subjects were asked to shut their eyes. The visual areas seem to be activated by the active somatosensory areas by cross-modal interaction, because the same objects are represented differently in different sensory cortical areas. These multiple representations seem to be linked cross-modally. Significant activation was observed in medial and lateral motor cortical areas but not in primary motor cortex. This finding confirms the role of the motor cortex to drive exploratory movements of haptic object recognition. It further demonstrates that these movements are complex movements that can only be driven by higher order motor cortical areas.

In this section, we have seen how the basic primitives of the somatic sense are combined to construct an image of the somatic object. We have also argued that the somatic object recognition is actually haptic object recognition since it does not passively depend on the somatosensory received bottom-up, but depends extensively on the active movements driven by the motor cortex. We have also briefly described and tried to rationalize the neural substrates of haptic object recognition as revealed by functional imaging studies. What does the somatosensory system do beyond haptic object recognition? This question will be discussed in the following section.

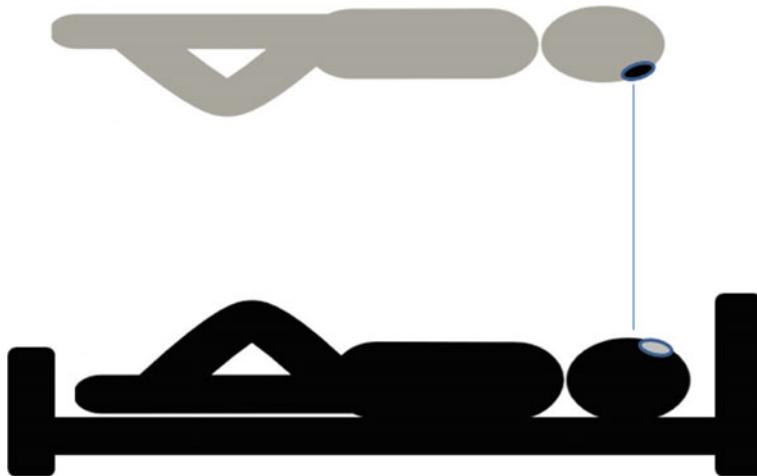
## Constructing the Body Image

A challenge that the somatosensory system addresses every moment of our existence is to represent not just the diverse array of 3D objects that we touch and manipulate in the world every day but to construct a living image of the most salient object of them all—our own body. We believe we are proud owners of a body, but what we possess in reality is only an image of it. The brain draws upon and integrates the streaming multisensory information from our body and constructs this living, unified image known as the body image. It is not just comprised of the visual appearance as instructed by the visual system; it also includes the feel of it as informed by the somatosensory system. It is through this image that we understand the spatial extent of our corporeal selves, our weight as it is borne by our muscles as we stand up, our mass or moment of inertia indicated by our postural systems as make rapid turns, etc. The integrated construct of all these different sources of information is called the body image.

The expression body image was first coined by Austrian neurologist and psychoanalyst Paul Schilder who used it not just in the low level, somatic sense as we have done above, but even in social and cultural senses. The concept of body image is used in a variety of disciplines including psychology, medicine, psychiatry, philosophy, and cultural studies. The body image consists of not just physical (mass, weight, etc.) or mensurational (height, width, etc.) properties but also its aesthetic and sexual qualities that dominate social and cultural attitudes toward the body. Thus, body image is not just a result of a faithful reconstruction of the information streaming from the sensory shop floors of the body. It includes a profound value judgement that depends on complex social and cultural factors. Distortions in this image, and the associated value judgement, lead not just to social maladjustment. Body image distortions can precipitate most dramatic clinical conditions like eating disorders, obesity, depression, and other forms of mental illness. In this section, let us consider some of these prominent distortions of body image and learn from these disorders how the body image operates under normal conditions.

## The Out-of-Body Experience and the Body Image

One of the most radical distortions of the body image is exemplified perhaps by the phenomenon of Out-of-Body Experience (OBE). It is a transient condition in which the subject reports “leaving” the body and taking an unusual spatial perspective that is not centered in the body. In the classic vanilla OBE, the subject’s body is in a supine position while the reports hovering on top of his/her body looking down on the body (Fig. 8.9). Innumerable variations, some intriguing and some simply bizarre, have been reported through the history. Unregulated reporting allowed a free play of imagination, blocking understanding and eliminating objectivity. The vacuum of comprehension was soon filled by religious interpretations in terms of existence of



**Fig. 8.9** A schematic depicting classical out-of-body experience

soul and a life beyond the body. However, experimental work done over the last few decades seemed to demystify OBE, demonstrating after all that the phenomenon has firm neurobiological origins.

The history of OBEs goes back more than a century ago to the late nineteenth century. One of the earliest reports of OBE was from a French otologist named Pierre Bonnier. In 1905, he described a patient who reported that he felt two forms of self: one of them still and other looking outward; then the two forms approached each other and merged. In a more outlandish version, a person who claimed to go off on nightly promenades outside his body was able to read, on one of his wanderings, a post-it note put up by his friend on his window at a faraway place. There were also reported cases in which accident victims took advantage of OBE, briefly abandoned their mutilated bodies on the highways, and rushed for help.

A scientific approach to OBE has begun to emerge in the late '80s. There was an approach from psychology that proposed that OBE arises when there is a breakdown of cognitive systems that construct a model of the "reality" and the body within that reality. There were also inroads into the OBE phenomenon from philosophy. Philosopher Thomas Metzinger distinguishes between the notion of pure self-hood and the body or the "self-model". He introduces the concept of Phenomenological Self-Model (PSM) to represent the pure self-hood distinct from the self-model associated primarily with the body. He defines the OBE as belonging to a class of self-states in which there is a single PSM but multiple self-models.

Increasing number of reports of OBE in both healthy and subjects with neurological conditions paved way to a neurobiological approach to understanding OBE. OBEs were reported in patients with epilepsy, migraines, post-traumatic brain disorder, and brain tumors. Here is a description by neurologist V. Lunn of an OBE experienced by a healthy subject:

Suddenly it was as if he saw himself in the bed in front of him. He felt as if he were at the other end of the room, as if he were floating in space below the ceiling in the corner facing the bed from where he could observe his own body in the bed... he saw his own completely immobile body in the bed; the eyes were closed.

There is an allied class of hallucinations in which the body is seen not just in a supine position as seen from above, but in other postures seen from other perspectives. (Note the term "hallucinations" used in neurological accounts of OBE, as opposed to "phenomenal self-model" used in a philosophical account. Philosophy seems to be willing to treat OBEs as one of the many ways of the phenomenological experience of world and self, whereas neurology treats OBEs *a priori* as hallucinations.) One such a hallucination is termed autoscopic hallucination. In this type of hallucination, the subject reports seeing a double of oneself in the extrapersonal space but, unlike OBE, does not have feeling of leaving one's own body. In a third class of phenomena, termed heautoscopy, which is midway between OBE and autoscopic hallucination, the subject reports seeing a double but is unable to decide whether he/she has left his/her body. In an interesting case of this kind described by Olaf Blanke and colleagues, the patient reports seeing herself from behind, a perspective that is normally denied to the rest of us lest by use of a peculiar arrangement of mirrors.

[The patient] awakens from sleep and has the immediate impression as if she were seeing herself from behind herself. She felt as if she were 'standing at the foot of my bed and looking down at myself.' Yet ... the patient also has the impression to 'see' from her physical visuo-spatial perspective, which looked at the wall immediately in front of her. Asked at which of these two positions she thinks herself to be, she answered that 'I am at both positions at the same time.' She did not have the feeling of being out of her body...

In more extreme cases, the subjects report seeing the world from two simultaneous or alternating vantage points. The resemblance of this phenomenon to alternating visual percept seen in case of binocular rivalry is striking. What is common to all the above phenomena is the fact of seeing the body from an extracorporeal position.

The fact that OBE is often experienced in neurological subjects only a general relation between neural pathology. A further insight in this matter was provided by the fact that the OBEs are linked to vestibular impairment. The vestibular system is the brain system that controls the sense of balance. Although, the semicircular canals of the inner ear that transduce the angular accelerations of the head supply part of the sensory information required for estimating balance, a more crucial source of balance is the proprioceptive information that comes from the entire musculo-skeletal system. Impaired vestibular sensations include sensations of elevation or floating or 180° inversion of the body. Such pathological balance-related sensations are reported by healthy subjects in the zero-gravity conditions of space missions, or low-gravity conditions of a parabolic flight.

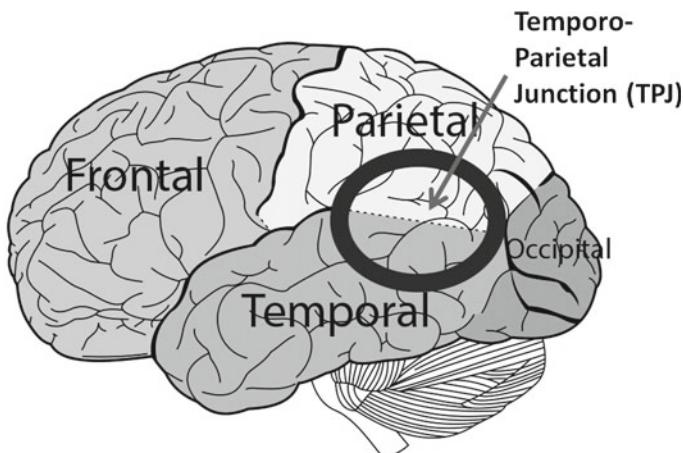
A more direct evidence supporting a neurobiological basis of OBEs came from the stimulation experiments of Olaf Blanke and colleagues, who found that OBE can be induced by electrical stimulation of the right Temporo-Parietal Junction (TPJ). In one such a simulation study, the patient reported OBE when the stimulation was given as the patient was looking straight ahead while fixating on any target. But when

the stimulation was given while the patient was looking at her stretched hands or legs, it produced an uncanny sensation of shortening of the limbs. When the limbs were bent at the elbow or the knee, stimulation produced a sensation of limb movement. These data present strong evidence that altered brain activity can produce illusions like OBEs or perhaps related phenomena such as autoscopic hallucinations.

OBE is a case of a radical distortion in self-image. If phantom limb phenomenon, described in Chap. 6, represents a distortion of the experienced image of a body part, OBE represents a distortion of such nature extended to the entire body so much so that it distorts not just how the body is felt but also how it is perceived from a visuo-spatial perspective. Considering the role of TPJ and surrounding areas in integrating the information about the body arising out of multiple sensory modalities, the contributions of these areas to OBE are not surprising.

The primary attractive feature of TPJ is its strategic location in the posterior cortex (Fig. 8.10). Firstly, it is located on the border between temporal and parietal cortex on the upper bank of the sylvian fissure. Therefore, it is well positioned to receive inputs from the three sensory modalities: somatosensory, auditory, and visual, while TPJ may not receive direct inputs from the sensory cortices of the above three modalities, it probably receives indirect inputs from the heteromodal cortical areas like the superior temporal sulcus. One fMRI-based study showed activation of TPJ and a nearby cortical area called the right posterior superior temporal cortex under conditions where there was a conflict between the visual and proprioceptive feedback from the hand. The study, therefore, indicates that TPJ has a role in integrating the visual and proprioceptive representations of body parts.

The role of TPJ in construction not just the body image, but the representations of proximal space itself, is brought to fore in the intriguing neurological condition of visuo-spatial hemineglect. Patients suffering from this condition tend to neglect the left half of the space and their body. For example, they may not notice objects placed



**Fig. 8.10** The temporo-parietal junction

to their left. The neglect is also manifested in terms of their body image: they might forget to wear their shirt on the left side. In a dramatic demonstration of the spatial deficits involved in hemineglect, Italian researchers Bisiach and Luzzatti, who lived in the city of Milan, asked their patients to imagine viewing the Piazza del Duomo, a famous monument, from the vantage point of a cathedral located at the end of the street. The patients were next asked to imagine viewing the cathedral from the Piazza del Duomo. The patients were asked to report what would be seeing from both vantage points. In both cases, the patients omitted structures located on the left side of the street with respect to their current vantage point. The experiment showed profound deficits induced by the hemineglect which affects not only visible spatial world but also memorized spaces. Hemineglect patients primarily have lesions, typically in the right hemisphere, in the angular (ang) and supramarginal (smg) gyri of the Inferior Parietal Lobe (IPL), the Temporo-Parietal Junction (TPJ), and the Superior Temporal Gyrus (STG).

Another case of a radical distortion of body image, a case with which we began this chapter, is anorexia. We had noted that the sufferer of anorexia believes that there is a mismatch between her felt body imagine and the actual, objective image of her body. She tends to feel she is overweight, constantly haunted by fears of gaining weight. This misperception results in abnormal dieting patterns leading to starvation. Although anorexia patients frequently have those anxious encounters with the weighing machine, their real trouble seems to be not just with their weight but the size of their body, their body image. For example, anorexia patients showed high anxiety when they encountered slim models, and even when they were shown line sketches of slim women. There seems to be an inaccuracy in the body image that they constructed for themselves. The impairment is probably in the part of the brain that integrates the physiological feedback from their body and constructs that body image.

Functional imaging studies that sought to unravel the neural substrates of anorexia have unearthed an entire network that is responsible for constructing the body image. This network is comprised of the dorsolateral prefrontal cortex, the lateral occipito-temporal cortex, and areas in the parietal lobe. Within the parietal lobe, some studies have found reduced activation in the temporo-parietal junction; in the temporal lobe such reduced activation was found in superior temporal sulcus and the temporal pole. It is noteworthy that the temporo-parietal junction figures repeatedly in syndromes related to construction of body image.

Being strategically located near the highest points of sensory processing hierarchy in the posterior brain, the temporo-parietal junction seems to be ideally posited to process the sensory information from both the internal world (sometimes referred to as the interoception) and from the external world. Empowered by this information, the temporo-parietal junction seems to be able to distinguish between the body and the world, delineate the border between the two, and dynamically maintain the body image. Like with many other questions of neural substrates of a brain function, the temporo-parietal junction does not perform this task in isolation but in cooperation with other brain areas in the prefrontal and more proximally within the parietal and temporal lobes.

More advanced studies on the functions of the temporo-parietal junction show that the role of this brain area goes far beyond the maintenance of the body image. Constructing body image consists of distinguishing between the body and the rest of the world. The temporo-parietal junction seems to extend the same ability to discriminate between one's own body and the body of another, laying foundations to what is termed the Theory of the Mind. When we discriminate between the body of another and ours, we are fully aware that it is not simply another body, but embodies another mind similar to one's own. When interacting with people, we believe and assume that just as we have a mind endowed with a sense of self, the others too, who basically present themselves to us in terms of their body and the behaviors it expresses, possess an independent mind with its separate sense of self. This belief in the self-hood of the other is called the Theory of Mind. Functional imaging studies showed that the temporo-parietal junction and the network of brain areas involved in creation of the body image also found to be neural correlates of the Theory of Mind.

Since humans are social creatures, effective engagement with other people or a harmonious participation in the society depends on our understanding of other people. Our ability to understand the contents of the minds of others and predict their intentions becomes a foundation for creating perfect social relationships. Poor or flawed Theory of Mind is, therefore, seen in psychiatric conditions that are characterized by social maladjustment like schizophrenia and autism spectral disorder.

The temporo-parietal junction seems to be a crucial node in a network of brain regions that collaborate to generate a dynamic body image. But to say that such and such brain region performs such and such a function has a recurrent blunder in the history of neuroscience. A completely satisfactory theory of the role of temporo-parietal junction and the associated areas in creating a body image and a sense of self can only originate from a detailed computational network model that takes the low-level sensory data as the input and demonstrates how the body image can be constructed by integrating such data over a complex hierarchical and recurrent structure. The model must be hierarchical because the integration must be performed over many stages, combining multimodal sensory sources hierarchically; it must be recurrent because it must involve the recurrent loop of sensory motor system. The self is something that must be constantly constructed and confirmed by the rich repertoire of interactions between the body and the world. By such interactions, the brain constantly redraws the borders between the body and world. Out of these interactions the neural self is born.

## References

- Ambedkar, B. R. (1969). *The untouchables* (2nd ed., pp. 1, 26). Shravasti, Gonda, U.P.: Bharatiya Boudha Shiksha Parishad.
- Aristotle. (1976). *Ethics* [Nicomachean ethics] (J. A. K. Thomson, Trans.). London: Penguin.
- Aristotle. (1986). *De Anima* [On the soul] (H. Lawson-Tancred, Trans.). London: Penguin.
- Aurobindo, S. (1998). *The upanishads: Texts, translations and commentaries*. Puducherry: Sri Aurobindo Ashram Press.

- Blanke, O., & Arzy, S. (2005). The out-of-body experience: Disturbed self-processing at the temporo-parietal junction. *The Neuroscientist*, 11(1), 16–24.
- Corsini, D. A., & Pick, H. L. (1969). The effect of texture on tactually perceived length. *Perception and Psychophysics*, 5(6), 352–356.
- Field, T. (2014). *Touch*. Cambridge: MIT Press.
- Fisher, J. D., Rytting, M., & Heslin, R. (1976). Hands touching hands: Affective and evaluative effects of an interpersonal touch. *Sociometry*, 39, 416–421.
- Gardner, E. P., & Kandel, E. R. Touch. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), (2000). *Principles of neural science* (Vol. 4, Chapter 23). New York: McGraw-Hill.
- Gardner, E. P., & Martin, J. H. (2000). The coding of sensory information. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (Vol. 4, Chapter 21). New York: McGraw-Hill.
- Gardner, E. P., Martin, J. H., & Jessel, T. M. (2000) The bodily senses. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (Vol. 4, Chapter 22). New York: McGraw-Hill.
- Gray, L., Watt, L., & Blass, E. M. (2000). Skin-to-skin contact is analgesic in healthy newborns. *Pediatrics*, 105, 14–20.
- Green, B. G. (1993). Heat as a factor in the perception of taste, smell and oral sensation. In I. Research & B. Marriott (Eds.), *Nutritional needs in hot environments: Applications for military personnel in field operations* (pp. 173–186). National Academy Press, Washington D.C.
- Hall, E. T. (1963). A system for the notation of proxemic behavior. *American Anthropologist*, 65(5), 1003–1026.
- Hall, E. T. (1966). *The hidden dimension*. New York: Anchor Books. ISBN 0-385-08476-5.
- Harlow, H. F. (1958). The nature of love. *American Psychologist*, 13(12), 673.
- Henley, N. M. (1973). Status and sex: Some touching observations. *Bulletin of the Psychonomic Society*, 2, 91–93.
- Henley, N. M. (1977). *Body politics: Power, sex, and nonverbal communication*. Englewood Cliffs, NJ: Prentice Hall.
- Hertenstein, M. J., Verkamp, J. M., Kerestes, A. M., & Holmes, R. M. (2006). The communicative functions of touch in humans, nonhuman primates, and rats: A review and synthesis of the empirical research. *Genetic, Social, and General Psychology Monographs*, 132(1), 5–94.
- Heslin, R., Nguyen, T. D., & Nguyen, M. L. (1983). Meaning of touch: The case of touch from a stranger or same sex person. *Journal of Nonverbal Behavior*, 7, 147–157.
- Hirose, K. (2009). The richness of touch: The paradoxical meanings of disability in Japanese culture. *The East Asian Library Journal*, 13(2), 59–85.
- Johnson, G. A. (2002). Touch, tactility and the reception of sculpture in early modern Italy. In C. Wilde & P. Smith (Eds.), *A companion to art theory* (pp. 61–74). Oxford: Blackwell.
- Klatzky, R. L., & Lederman, S. J. (2011). Haptic object perception: Spatial dimensionality and relation to vision. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1581), 3097–3105.
- Lunn, V. (1970). Autoscopic phenomena. *Acta Psychiatrica Scandinavica*, 46(Suppl 219), 118–125.
- Major, B. (1981). Gender patterns in touching behavior. In C. Mayo & N. M. Henley (Eds.), *Gender and nonverbal behavior* (pp. 15–37). New York: Springer.
- Metzinger, T. (2005). Out-of-body experiences as the origin of the concept of a ‘soul’. *Mind and Matter*, 3(1), 57–84.
- Montagu, A., & Montague, A. (1971). *Touching: The human significance of the skin* (p. 292). New York: Columbia University Press.
- Morris, D. (1971). *Intimate behavior*. New York: Random House.
- Nico, D., Daprati, E., Nighoghossian, N., Carrier, E., Duhamel, J. R., & Sirigu, A. (2010). The role of the right parietal lobe in anorexia nervosa. *Psychological Medicine*, 40(9), 1531–1539.
- Plaisier, M. A., Tiest, W. M. B., & Kappers, A. M. (2008). Haptic pop-out in a hand sweep. *Acta Psychologica*, 128(2), 368–377.

- Reed, C. L., Shoham, S., & Halgren, E. (2004). Neural substrates of tactile object recognition: An fMRI study. *Human Brain Mapping*, 21(4), 236–246.
- Sarukkai, S. (2009). Phenomenology of untouchability. *Economic and Political Weekly*, 39–48.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind”. *NeuroImage*, 19(4), 1835–1842.
- Schilder, P. (2013). *The image and appearance of the human body*. UK: Routledge.
- Thayer, S. (1986). History and strategies of research on social touch. *Journal of Nonverbal Behavior*, 10(1), 12–28.
- Treisman, A. (1998a). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 353(1373), 1295–1306.
- Treisman, A. (1998b). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology Section A*, 40(2), 201–237.

# Chapter 9

## Life in Motion



*Great ideas originate in the muscles.*

—Thomas A. Edison.

### Primeval Motion

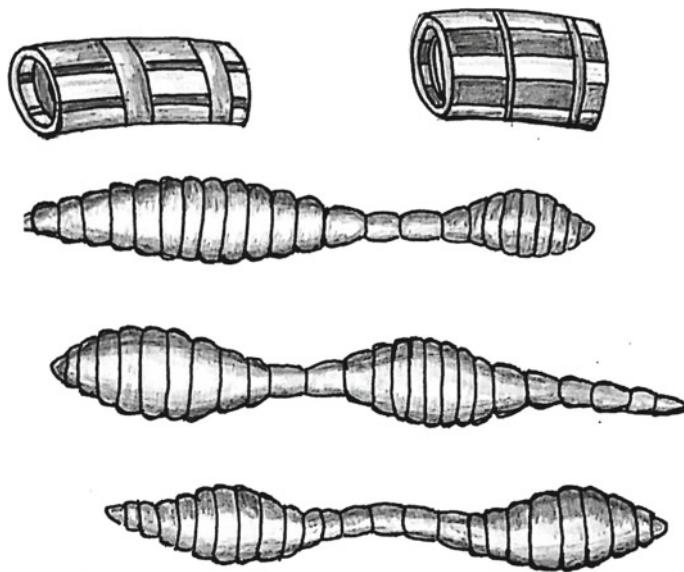
Motion is generally believed to be a defining feature of life. By the power of motion an organism hunts down its prey, flees its predators, shifts to superior habitats and moves from where it is to where it wants to be. (Familiar stationary life forms—the plant life that we see around us—do not have the same advantage.) Like the opposable thumb, or encephalization, the ability to move, to proactively propel itself and plough through a resisting world marks a great milestone of evolution.

But some of the earliest organisms did not move, like the sponge, for example. They have no limbs or gills or any other special appendages with which they can accomplish active motion. Sponges live in water bodies, both freshwater and marine. For the most part of their lives they remain tethered to fixed surfaces—they are sessile. Therefore, unlike other self-respecting critters that can move and actively seek out food, sponges get food by passive solicitation. One can image a typical sponge as a large, porous rubber vat—surrounding water flows into the vat through the pores in its wall, collects in a central cavity, from where it is pumped upwards to be ejected out of an orifice. As the water flows through the pores of the wall, food particles present in the water gets trapped inside the walls, and subsequently absorbed. Sponge cells also absorb dissolved oxygen in the water and excrete carbondioxide and other waste material into the water flowing through its pores. Some sponges do move on the seabed, albeit painfully slowly, at a pace of about 1–4 mm per day. They achieve this humble motion by amoeba-like movements of certain specialized cells called pinacocytes. Juvenile sponges live a free, untethered life, drifting freely, pushed by

the ocean currents. However, once they arrive at adulthood and maturity they settle down, anchoring themselves to a secure surface.

An interesting next step from the sponge is a creature that shows beginnings of active propulsion—the jellyfish. With the appearance of some sort of a soft, translucent umbrella with slow, ghost-like movements, jellyfish attract curious visitors to aquaria all over the world. The umbrella-shaped central body, known as the “bell”—has an internal water-filled cavity. The bell is also surrounded by a muscle which can be controlled by the diffusive nerve net spread out all over its body. Contraction of this muscle results in contraction of the entire bell, thereby expelling the water contained in the chamber via a mouth located in its underbelly. The resulting hydrostatic propulsion moves the creature forward. This propulsive effort of the jellyfish had attracted close scrutiny of scientists who found that jellyfish adjust their propulsive movements cleverly to efficiently utilize energy for propulsion. Once the jellyfish contracts its body and expels the water inside, it pauses briefly before contracting again. In the meantime, the “hole” created by the expelled water in the wake of the animal, is refilled by the surrounding water, which creates further propulsion. Based on his studies of jellyfish propulsion, Shashank Priya, a professor of mechanical engineering at Virginia Tech University, concluded that this special propulsion rhythm enables the creature to move 30% farther in every stroke cycle. These learnings from jellyfish swimming have inspired creation of a life-like, autonomous, robotic jellyfish for the US navy.

While the jellyfish hauled itself forward with a single ring of a muscle, there is another creature—the lowly earthworm—that wriggles with the help of a row of ring muscles and some more. Figure 9.1 shows a simple schematic of the muscles of an earthworm. There are two sets of them—the rings that surround it encircling its slithering body, and the longitudinal muscles that run along the length of its body. Then there are these slender bristles called the *setae* with which it temporarily tethers itself to a surface with which it is in contact. Another important aspect of an earthworm that must be remembered to understand its motion is its body itself which is like a fluid-filled long balloon. Such bodies are known as hydrostatic skeletons. An earthworm is not endowed with a skeleton. It is soft and slimy all over. Its fluid-filled interior therefore creates the necessary stiffness or turgidity necessary for the body to wriggle around in the world. If you squeeze such a balloon, since fluids are generally incompressible, the body simply elongates in a direction permitted by the squeeze. If you squeeze, for example, its posterior, the anterior portion elongates. Now consider the following sequence of contractions of the earthworm’s muscles that haul its body forward gently in stages. First, the ring muscles in the posterior part of the body contract; the ring muscles of the anterior muscles also contract but to a lesser extent; the longitudinal muscles in the anterior part relax. Since the posterior muscles are in a state of contraction, the fluid in the body moves to the anterior part. In the anterior part, since the ring muscles are still in a state of contraction, its body cannot expand radially, but only stretches out longitudinally, pushing the creature forward. At the same time, the setae on the posterior side are stretched out so that they anchor the posterior part of the body to the ground, preventing it from slipping backward as a reaction to the forward thrust generated in the anterior part of the body. Through such



**Fig. 9.1** The writhing, propulsive movements of an earthworm. The top two figures show the arrangement of longitudinal and ring muscles

sequential contractions pulsing along the length of its body, an earthworm softly and quietly ploughs through the soil.

The idea of a hydrostatic skeleton and the movements controlled by its twin systems of radial and axial muscle systems, paves way to the discussion of a similar systems found at several levels of evolutionary hierarchy, capable of a great diversity of movements. In a hydrostatic skeleton, there is a muscle that surrounds a fluid-filled cavity. But such a cavity is dispensable since the muscle is itself a fluid-filled organ and therefore primarily incompressible. Therefore, when a muscle shortens, it becomes fatter; when it elongates it becomes leaner. The resulting rope-like muscular structure that can be used for grasping, lifting, probing, digging, crushing, snapping, and other movements in a variety of species is known as a muscular hydrostat. The muscles in the muscular hydrostat are typically oriented in three different directions: parallel and perpendicular to the long axis, as in case of the earthworm, and oblique to the long axis, in addition. The elephant's trunk or proboscide, octopus' arms, nautilus' tentacles are all excellent examples of muscular hydrostat. Come to think of it, we do not need to stray too far afield to locate a good muscular hydrostat—our tongue is perhaps the most eloquent example. The rich and complex movements of the tongue involved in speech articulation and mastication of food is a private and convincing demonstration of what a muscular hydrostat is capable of.

So far the examples of muscular action that we have seen consisted of soft bodies and fluid-filled, floppy structures. But muscular action without the framework of the skeletal system, that bony gridwork over which the muscles act, in order to

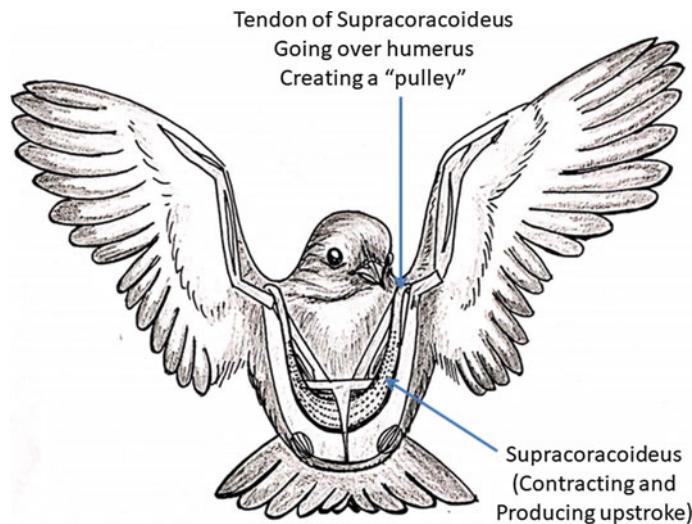
produce a complex interplay of pulling, pushing, twisting, and turning forces, is an incomplete story. A fine example of a system wherein the force-generating but soft muscles, interacting with a passive but rigid skeletal system, that enabled evolving biomechanics to take to the skies is the avian musculoskeletal system. Unlike artificial flying devices that use rotors, propellers, and jets of all sorts to produce thrust, birds use flapping wings to produce lift.

Flight is a high-power activity since the bird has to lift its entire weight working against gravity, and push it forward working against air drag. For this purpose, birds typically need muscles that are so massive they occupy nearly a third of their body mass. To understand the muscular requirements of avian flight, let us compare the flapping movements of wings of a bird with movements of human arm. (In fact, legend has it that in ancient Greece, Daedalus and Icarus, a father and son duo, tied wooden wings on their arms, with feathers stuck in the wings using wax, and tried to fly in a bid to escape from an island prison. Myth aside, the forces necessary to work the artificial wings would be superhuman, bordering on the impossible.) Imagine moving your arms, initially stretched out sideways, now stretched out straight in front of you. This movement is generated by breast muscles called pectoralis major that originate along the breastbone or sternum and insert near the head of the upper arm bone called the humerus. This movement is analogous to the downstroke of a bird's wing.

Now imagine the opposite movement of the arms, from the front to the sides again. The same muscles cannot perform this opposite movement since muscles can only pull, and not push. Therefore, to produce the opposite movement, the body engages a different set of muscles, referred to as the antagonist muscles, in contrast to the first set known as the agonists. Thus, the agonist and antagonist muscles are dual sets of muscles that produce opposite movements. This opposite movement from the front, back to the sides, is produced by the action of deltoids, placed on the top of the shoulder. You can feel this muscle bulging slightly when you try to lift your arm.

In the avian case, it is the downward thrust that produces lift and therefore must be much stronger than the upward thrust. Both the downward and upward forces generated by the avian muscles, relative to its body weight, are much greater than those generated (or required) in case of humans. For this purpose, avian musculoskeletal system had undergone some smart adaptations. First of these is the presence of a vertical keel, a strong ridge-like extension to the sternum, that greatly expands the surface area for muscle attachment. The large keel supports the strong action of the pectoralis major, as it pulls the wing down, generating lift.

If the pectoralis major, the muscle that produces lift, is in the underbelly, the muscles that produce the opposite movement, lifting the wings up, must be logically located on the back. But this second muscle, *supracoracoideus*, also connected to the keel, lifts the wings up, since it connects to the top of the humerus by way of a pulley. Using a simple pulley we know that we can lift a weight, by actually pulling a rope down: a pulley helps us to change direction of the applied force. Likewise the tendons of supracoracoideus going over a special groove of humerus, changes the direction of the applied force (Fig. 9.2). By this arrangement, both the heavy muscle



**Fig. 9.2** The supracoracoideus muscle located in the breast of the bird goes over the groove of humerus. Contraction of this muscle changes the direction of the force and produces upswing of the wings

*masses, used for raising and lowering wings, are located in the underbelly of the bird. Such weight distribution gives the bird a greater aerodynamic stability.*

In the examples that we have visited so far, locomotion is ultimately driven by one or more muscles. Muscles are basically flexible, rubber-like strands that can only pull. They contract and exert force on the ends to which they are tethered. In an artificial engine, there might be a great diversity of mechanisms driving motion—burning chambers with beating pistons, jets of fluid or plasma, whirring rotors driven by force fields and so on. But life forms are simply driven by these contracting rubber strands—complex systems of them, pinned in intricate ways to the bony skeletons of the bodies, where such skeletons are present, or to their floppy counterparts, the hydrostatic skeletons, contracting in complex sequences driven by the commands from the host's nervous system, centralized or decentralized. All the scampering and scurrying, the flapping and the fluttering, the strutting and soaring, the hopping and galloping of life is but a manifestation of the biological puppetry of pulsing muscles, orchestrated by the neural puppeteer.

## Strands that Pull

Although hidden under the skin, the way other internal organs are, the presence of muscles can be quite easily seen, or felt, even as the body goes through its daily motions. A young preteen, trying to flaunt his/her meager and tentative strength to

the world, would flex the arm at the elbow, and display a muscle swelling like an ocean tide. One of the most popular muscles of the musculoskeletal system (if there is only a single muscle in your body you knew by its name, it would probably be it), the biceps, as any other muscle, swells in the middle when it contracts. Since the muscle is a fluid-filled structure, and since fluids are typically incompressible, a shortening muscle has to swell to conserve volume. In this simple act of muscle flexing, a gesture in its more dramatic forms can have sleeve-splitting consequences, lie the quaint origins of the word “muscle.” The root word “mus” refers to a mouse in Greek, and “musculus” specifically refers to a little mouse. Our imaginative townsfolk of antiquity visualized some sort of a subcutaneous mouse that presses against the skin whenever a person flexes the biceps.

Some of the earliest documented comments on the nature of the muscle, though quite misleading and misdirected, come from the writings of Greek philosopher Aristotle. In his voluminous *De Motu Animalium* (The Movement of Animals), this is what he wrote about muscles.

Now the functions of movement are pushing and pulling, so the tool of movement has to be capable of expanding and contracting. And this is just the nature of the *pneuma*. For it contracts and expands without constraint, and is able to pull and push for the same reasons; and it has weight by comparison with the fiery and lightness by comparison with the opposite. Whatever is going to impart motion without undergoing alternation must be of this kind....

Aristotle seems to have mastered the art of saying a lot while actually saying nothing. The above “account” of muscular action hardly holds any water, if we go by the modern standards of a scientific explanation. The basic issue with the above account, as with the general approach to physical phenomena by the antiquity, is the attempt to explain physical occurrences in terms of the hypothetical *five elements*. In many ancient cultures it was believed that the physical world and the interactions among physical objects are describable in terms of transformations among five elements dubbed as earth, water, fire, air and ether. In the above paragraph, Aristotle informs us that the muscle is heavier than “fire” and lighter than “earth.” It is impossible to confirm or reject such a statement since the notion of the elements is vague and cannot be rooted empirically in day-to-day experience. He further says, and this is where he lends himself to a more solid criticism based on physical arguments, that since movement consists of “pushing and pulling,” any instrument of movement (here, muscle) must be capable of “expanding and contracting.” But such a statement is broken at multiple places. First of all, muscles are not capable of “expanding and contracting” as a whole; if they contract in one part (say, lengthwise), they expand in another (say, girthwise). Second, they can only pull and not push. “Pushing” is achieved essentially by the pulling action of a different muscle.

A prominent figure who lived in the early renaissance period in Europe, and was involved in some serious myth-busting in the area of medicine and physiology was Andreas Vesalius. At the age of 28, he published his prodigious studies on human anatomy, righting a large number of anatomical wrongs perpetrated by his predecessors like Aristotle and Galen. Vesalius writes about muscular action as follows:

...I am persuaded that the flesh of muscles, which is different from everything else in the whole body, is the chief agent, by aid of which (the nerves, the messengers of the animal spirits not being wanting) the muscle becomes thicker, shortens and gathers itself together, and so draws itself and moves the part to which it is attached, and by the help of which it again relaxes and extends, and so let us go again the part which it had so drawn.

A positive feature of Vesalius' description of muscular action is that it is free from the five elements and other esoteric admixture. His brief commentary contains two key ideas: (1) muscles shorten (and therefore thicken) and thereby exert force on its two ends, and (2) this shortening is driven by neural input to the muscle. Although subsequent developments in muscle physiology have greatly fine-tuned this picture, the essentials remain.

Let us try to describe the above neutrally driven shortening of the muscle in more modern engineering terms as follows. We have earlier compared muscles with rubber bands. Let us imagine a rubber band with a pan that is practically weightless, attached at the lower end and suspended from a fixed ceiling. Let the length of the rubber band now be  $L_0$ , a quantity known as the *resting length*. Imagine now that you are adding a weight  $W$  to the pan; the increased weight of the pan pulls the rubber band down, elongating it to the new length,  $L = L_0 + a$ . Now if you add two weights of value  $W$ , the new length of the rubber band will be  $L = L_0 + 2a$ . You can easily verify that the increase in length is proportional to the increase in weight. Now if you gradually reverse the process, and remove one of the weights  $W$ , bringing the weight to just  $W$ , the rubber band contracts again to the older value of  $L_0 + a$ . If you remove the remaining weight  $W$  also, leaving the empty weightless pan, the rubber band returns to its original length of  $L_0$ . Thus, we can make a general summary of the above experiment as follows. The increased length ( $\Delta L$ ) of the rubber band will be found to be proportional to the increased weight ( $\Delta W$ ) acting on the rubber band.

$$\Delta W \propto \Delta L$$

Or

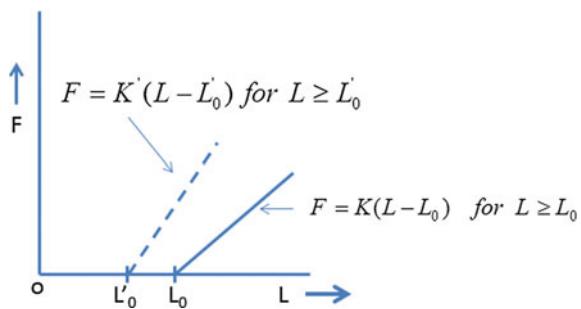
$$\Delta W \propto (L - L_0)$$

By inserting a proportionality constant,  $K$ , we have a simple equation,

$$\Delta W = K(L - L_0)$$

Note that the above equation is valid only when the rubber band is stretched. A regular run-of-the-mill rubber band can only be stretched and cannot be compressed to a length less than its resting length. (But a muscle is different, as we will see in a moment.) Only when the rubber band is stretched it generates a force,  $F$ . The situation where  $L$  is less than  $L_0$ , is impossible and in the equation above, force generated must be zero for that case. Therefore, we describe the rule relating force generated,  $F$ , and length of the rubber band,  $L$ , as follows:

**Fig. 9.3** Force versus length relation for a simple muscle model in resting conditions (solid line) and with neural activation (dashed line)



$$\begin{aligned} F &= K(L - L_0) \quad \text{for } L \geq L_0 \quad (\text{force-length equation}) \\ &= 0, \quad \text{otherwise} \end{aligned}$$

A graph that depicts the variation of  $F$  with respect to  $L$  is shown in Fig. 9.3 (solid line). Once  $L$  crosses a threshold,  $L_0$ , the force increases linearly; for  $L$  less than  $L_0$ ,  $F$  remains zero. The constant  $K$  is called stiffness. A stiffer rubber band stretches less for the same force, than a less stiff material.

The above pattern of variation of force  $F$  with respect to length  $L$  of a rubber band is a reasonable preliminary description of a muscle that is stretched passively.

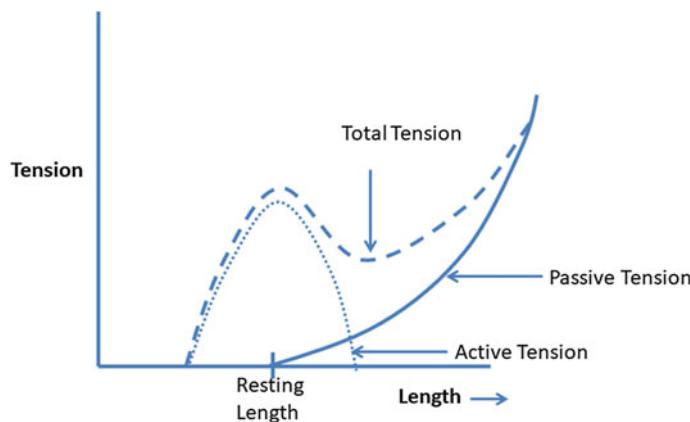
But several times in the preceding discussion, we spoke of the neural activity causing “contraction” of the muscle. We can make an easy alteration of *force-length equation* to describe the effect of neural activation on the muscle.

Now imagine a rubber band made of a special, magical material whose stiffness,  $K$  can be increased to  $K'$  and the resting length,  $L_0$ , can be decreased to  $L_0'$ , instantaneously. The *force-length-equation* above can now be written as

$$\begin{aligned} F &= K'(L - L_0') \quad \text{for } L \geq L_0' \\ &= 0, \quad \text{otherwise} \end{aligned}$$

If the new force  $F$  versus  $L$  relationship is plotted in the same graph, it looks something like in Fig. 9.3 (dashed line). The magical rubber band, which is now behaving like a muscle under neural activation, starts generating force even at shorter lengths, and the rate at which the force is generated is greater than for a passive muscle that does not receive neural activation.

What we just described is the behavior of an idealized rubber band posing as a muscle. The tension generated by a real muscle when stretched passively does not increase linearly with length. It shows a greater than linear increase with length, curving upwards as in Fig. 9.4. Now, when the muscle is activated by neural or electrical stimulation, it generates a greater tension than in the case of passive stretching, for the same length (Fig. 9.4). However, this greater force is seen only for a limited range of lengths of the muscle; for greater lengths neural activation and passive stretching generate nearly the same force. That is, for greater lengths, neural activation fails to



**Fig. 9.4** More realistic (than in Fig. 9.3) tension versus length relationship in a muscle

generate additional force. Thus the “active” force generated by neural stimulation has a bell-shaped profile. This bell-shaped active force versus length profile of a muscle has its roots in the intriguing molecular machinery of the muscle, a story that took a few centuries to completely unravel. We will discuss some of these details in the following section.

Thus we have seen that the force generated by a muscle depends on the actual length and the state of stimulation of the muscle. There is, however, another crucial factor that determines muscle force. Not just length, but the rate of change of length of the muscle also, determines the muscle force. This rate of change of length, intriguingly called “velocity” in muscle physiology literature, is defined as the rate of shortening of the muscle. A shortening muscle, therefore, by convention, is said to have a positive velocity, while an elongating muscle has a negative velocity. It turns out that force generated is much higher for negative velocity (elongating muscle) than for positive velocity (shortening muscle). This inverse relation between muscle velocity and force is expressed by a simple formula, known as the Hill equation, named after A. V. Hill, in recognition of his pioneering studies on muscle physiology.

The fact that elongating muscles generate a greater force than shortening ones, is brought to fore in a dramatic fashion by a demonstration that took place at the Royal Society in London. This demonstration, as described by A. V. Hill, consisted of two bicycles, facing each other, and connected together with a single long chain going over their cranks. Thus, if the pedals of one of the bicycles are pushed forward, the pedals of the other turn backwards. Two people mount the two bicycles. As the person on one of the bikes tries to pedal forward, the person on the other bike tries to resist the backward motion of the pedals in his/her bike. In this experiment, the first bicyclist—the one pedaling forward—was always driven to exhaustion very soon, while the second cyclist, who is merely resisting, won. In this cunning arrangement, the first cyclist tries to generate force by shortening his/her muscles, while the second

cyclist tries to produce force by elongating the same muscles. The second person generates greater forces, and prevails over the first one.

Finally a word on the idea of neural activation causing muscular contraction. Earliest studies in this area date back to the eighteenth century. The middle of the eighteen century saw the emergence of the field of “animal electricity.” That was a time when the effects of electricity on the human body were begun to be appreciated. The discovery of the action of electricity on muscular contraction happened serendipitously, in Luigi Galvani’s lab in Bologna, Italy. Galvani, who began his career as an anatomist, shifted his attention to the effect of electricity on animal tissues. As the story goes, one fine day Galvani’s assistant touched the sciatic nerve connected to a dead frog’s leg, with a charged metal rod. Sparks rushed from the charged rod and, passing through the nerve, activated the muscles of the leg; the dead leg moved as if life returned to it. These early experiments became the founding pillars of modern electrophysiology.

The simple experiments of Galvani that showed how an electrical stimulation applied to a muscle led to contraction of the muscle, begged the question of how an electrical signal can cause a mechanical effect. The engineer is familiar with this conversion of electrical energy into mechanical motion in an electric motor. But what exactly happens in a muscle? This was a question that took over two centuries to unravel. To understand the connection between electrical stimulation and mechanical contraction of the muscle, we must take a closer look at the microanatomy of the muscle, and consider the molecular players that connect electricity to mechanics in the narrow corridors of muscle tissue.

## The Innards of a Muscle

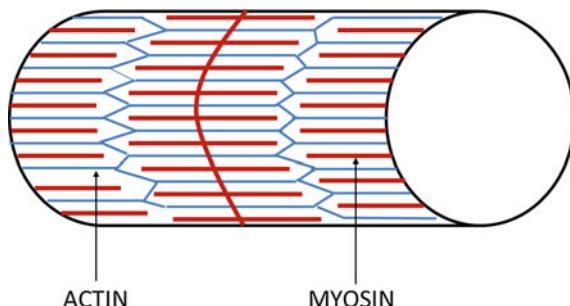
The theory that describes the microscopic machinery underlying muscular contraction was first published in 1954. The theory, intriguingly, was proposed simultaneously, dramatically published in the same issue of *Nature* (May, 1954), in two separate papers. The first paper by Andrew F. Huxley and Rolf Niedergerke was entitled: “Interference microscopy of living muscle fibres.” The second paper by Hugh Huxley and Jean Hanson was entitled: “Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation.” It is interesting that the lead authors of both the papers, totally unrelated to each other, shared the same last name—Huxley. Although the theory in its original form was proposed only about half a century ago, the events that led to that grand development can be traced back to the earliest microscopic observations of Antoine von Leeuwenhoek. Using his primitive optical microscope, Leeuwenhoek observed skeletal muscle and noticed that muscles are, first and foremost, bundles of fibers. These fibers that run along the length of the muscles are punctuated by certain band structures, some of which were even named by Leeuwenhoek himself. Improved methods of microscopy developed over the next couple of centuries have greatly elaborated the structure of the muscle at microscopic and molecular level, though there are still some outstanding questions.

According to the current understanding, the skeletal muscle is comprised of a bundle of parallel muscle fibers. Each fiber is a single cell or actually a chain of cells that amalgamated into a single longish strand. The fibers are typically a few centimeters long and about 10–100  $\mu\text{m}$  in diameter. The muscle fibers, being cells by nature, have a cell membrane called the sarcolemma. Inside the sarcolemma there are finer filaments called the myofibrils and a membranous network structure called the sarcoplasmic reticulum.

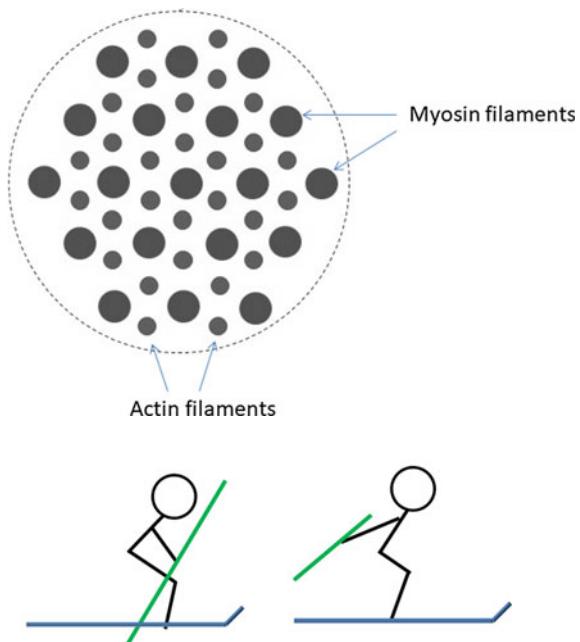
Just as a muscle is a bundle of muscle fibers, a myofibril is a bundle of two types of interleaving filaments that are essentially molecular chains, one thicker than the other. The thicker filaments are made up of a molecule called *myosin*, whereas the thinner filaments are comprised of *actin* (Fig. 9.5). Each myosin filament is surrounded by six actin filaments and each actin filament is in contact with three myosin filaments (Fig. 9.6). When a muscle changes its length, the actin and myosin filaments slide over each other, facilitating the length change. The actin filaments are attached at one end to a Z-line. Two sets of actin filaments and one set of myosin filaments, located between two adjacent Z-lines constitute a sarcomere. The actin and myosin filaments are connected by another system of molecular links generally described as cross-bridges. These cross-bridges, which are tethered to myosin filaments, repeatedly attach themselves to, pull, and detach themselves from the actin molecules. The action of the myosin filaments on the actin filaments, with the instrumentality of the cross-bridges, is analogous to that of a skier on the snow with the instrumentality of a ski pole. The skier digs his/her ski pole into the snow, drags the snow backwards, thereby propelling himself/herself forward, and pulls out the ski pole, only to repeat the sequence (Fig. 9.7). By an exquisite molecular ratcheting action, the cross-bridges pull the actin filaments inwards, thereby contracting the muscle.

The above-described molecular events that lead to contraction of the muscle are actually triggered by neural signals. In Chap. 1, we had learnt about Otto Loewi's classic experiment that demonstrated that neurotransmission is mediated by a chemical messenger. In that experiment it was shown that when the vagus nerve is stimulated, the beating heart to which the nerve is attached slows down, by the action of a chemical released from the nerve. This chemical, known as acetylcholine, acts on the cardiac muscle causing reduced force and frequency of contraction. Interestingly, even when a nerve innervating a skeletal muscle is stimulated, the same chemi-

**Fig. 9.5** Actin and myosin filaments in a myofibril



**Fig. 9.6** Interleaved arrangement of actin and myosin filaments as seen in a cross-section of a myofibril. Each myosin filament is surrounded by six actin filaments and each actin filament is in contact with three myosin filaments



**Fig. 9.7** Interactions between actin and myosin compared with the actions of a skier with a ski pole

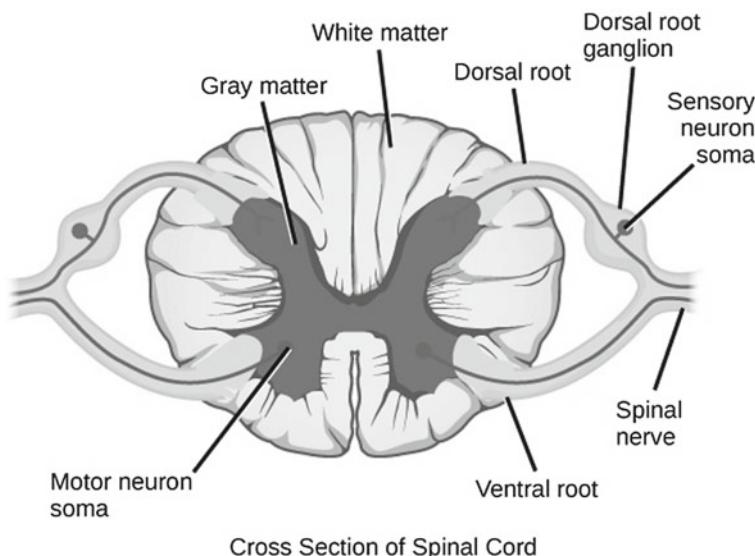
cal is released, but with a key difference. Acetylcholine *activates* skeletal muscle, causing it to contract and increase its force of contraction. When acetylcholine is released at the point of contact between the nerve and the muscle, a place known as the neuro-muscular junction, it acts on the large number of acetylcholine receptors (nearly 10,000 receptors/ $\mu\text{m}^2$ ). The activated receptors open  $\text{Ca}^{2+}$  channels, allowing a massive influx of calcium ions into the sarcoplasm, thereby increasing the  $\text{Ca}^{2+}$  concentration by nearly 100 times. It is this increased concentration of  $\text{Ca}^{2+}$  ions that enables the ratcheting action of cross-bridges, fuelled by the availability of ATP in the muscle fibers, resulting in muscular contraction.

## The Motor Unit

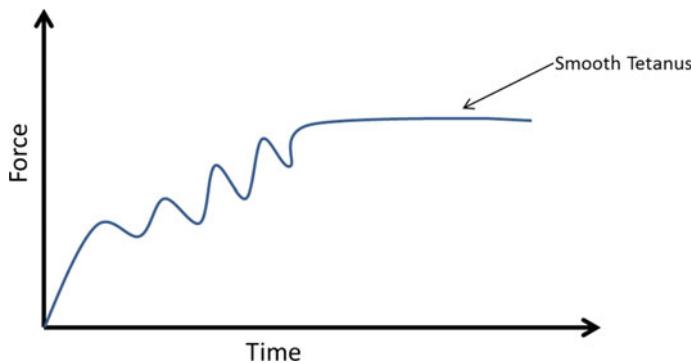
Thus when an action potential arriving at the end of an axon stimulates a muscle fiber, it initiates a complex sequence of molecular events inside the fiber that make the fiber produce a brief contractile force or *twitch*. These axons are part of a special class of neurons called the motor neurons, or more formally known as the alpha motor neurons, to distinguish them from a different class of motor neurons called the gamma motor neurons. Being directly responsible to stimulate skeletal muscle everywhere

in the body, the alpha motor neurons send axonal projections to the muscle fibers. Higher motor regions like the primary motor cortex produce effects on the skeletal muscle only by the intervening role of the alpha motor neurons of the spinal cord. The many brain regions that happen to have an influence on muscular action have to invariably funnel that influence through the alpha motor neurons. Therefore, Sir Charles Sherrington has famously dubbed the alpha motor neuron system as the “final common pathway” of motor action.

Clusters of alpha motor neurons found in the brain stem control the muscles of the head and the neck. These are the neurons that control our speech, facial expressions, movements of the eyes as they dart from target to more salient target, and so on. Then there are alpha motor neurons that control the muscles of the body below the neck. Movements of the hands, the postural movements of the trunk and the abdomen, and the movements of the legs are controlled by these neurons, which are located in the spinal cord. A section of the spinal cord reveals a whitish region surrounding a light brownish core, shaped like the wings of a butterfly (Fig. 9.8). This winged core is symmetric about the mid-sagittal plane (the vertical plane that cuts the body into two symmetric left-right halves). The wings, called the “horns,” extend both dorsally (backwards) and ventrally (forwards). The alpha motor neurons, located in the ventral horn send out their axons to the muscles of the same side of the body. These are some of the longest axons in the body; their proud neuronal owners are therefore some of the largest cells, in terms of spatial extent, in the body.



**Fig. 9.8** Cross-section of the spinal cord showing the dorsal and ventral roots, and the gray and white matters



**Fig. 9.9** Tetanus—a step-like force build-up pattern in a muscle

A single alpha motor neuron sends out its axonal projections to a large number of muscle fibers. Activation of an alpha motor neuron, therefore, produces simultaneous contraction in all the muscle fibers innervated by it. Thus the alpha motor neuron and all the muscle fibers innervated by it form some sort of a unit, conveniently named the “motor unit.” The number of fibers that are innervated by the alpha motor neuron, a quantity known as the innervation ratio, varies from muscle to muscle. In a small and delicate muscle like those that control the eye movements, the so-called extraocular muscles, each alpha motor neuron activates only three fibers. In a large muscle like the medial gastrocnemius, one of the calf muscles, the innervation ratio could be as high as 1900.

When a single alpha motor neuron fires, all the fibers innervated by it twitch at the same time. Thus motor units with larger innervation ratios generate a greater force when activated. On the same lines, activation of more motor units also results in a greater force generation. The above two ways of generating greater force of contraction in a muscle involve activating more units, analogous to the process of spatial summation at the soma of a neuron, where the greater the number of excitable dendritic inputs, the higher is the probability of the neuron’s activation. There is a second way of increasing muscle force using some sort of temporal summation. When a single action potential activates a muscle fiber, the force produced is a smooth wave that last for several tens of milliseconds, much longer than the millisecond long action potential that triggered the twitch. When a train of action potentials arrive at the muscle fiber in rapid succession, the fiber that had just begun to respond to a spike will not have time to relax before it is hit by the next spike. Therefore, the force produced builds on top of the previous wave, producing a step-like pattern of force build-up known as the *tetanus* (Fig. 9.9).

Thus by controlling the number of active motor neurons that innervate a muscle, or the firing rates of those neurons, the force generated in a muscle can be controlled. Compare the gentle burden of holding a pen in your hand, with carrying the dead load of large suitcase with the muscles of your hand, arm, and shoulder strained to the point of acute pain. An interesting question that was often asked in muscle

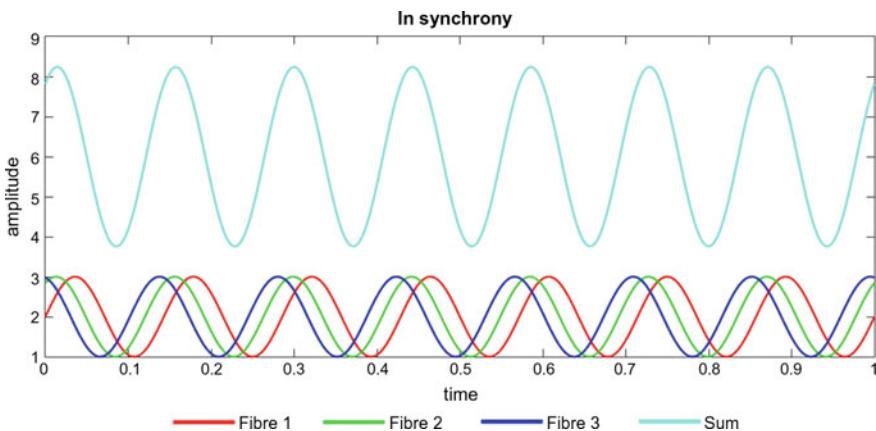
physiology, answered more or less satisfactorily is as follows. How does the brain know how many motor units (and which ones?) to activate in order to generate a certain force level in a given muscle?

Alpha motor neurons of the spinal cord receive their commands from higher motor centers in the midbrain and the motor cortex. It is generally thought that these commands do not specify which muscle fibers are to be activated in a given muscle. The higher commands to the spinal cord perhaps specify which muscle must be activated and at what level. The task of determining the exact selection of the motor neurons (and their associated motor fibers) is left to be elaborated by the foot soldiers of the cord—the alpha motor neurons and the related spinal circuitry of which they are a part. The strategy of this elaboration is often described using the *size principle*.

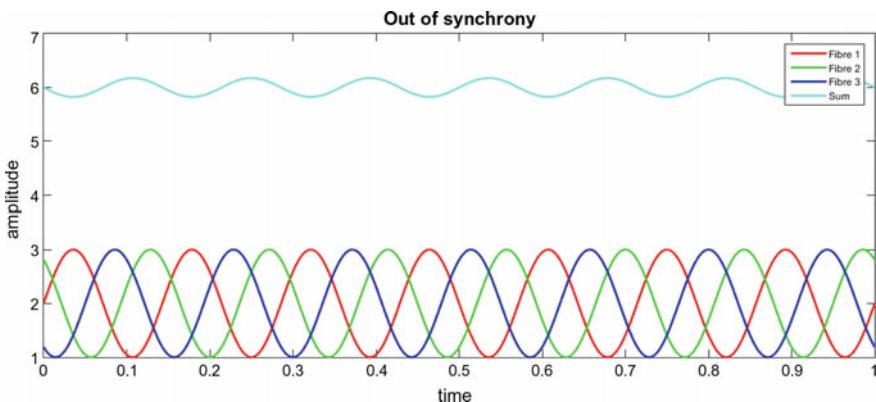
According to this principle, each alpha motor neuron has a threshold for activation. Furthermore, these neurons have a continuously varying thresholds of activation. To make arguments facile, let us consider a situation in which the relevant numerical quantities are greatly simplified. The arguments, however, can be easily extended to more biologically realistic descriptions. Assume that there are 100 neurons with their threshold of activation increasing systematically from 1 to 100. Now let us imagine that the command from the higher center, that only specifies the desired overall force,  $F$ , but does not decide the activation of individual motor units, numbered between 1 and 100. This number is given equally to all the alpha motor neurons. Although all the alpha motor neurons receive the same input,  $F$ , a neuron is activated only if the input exceeds its intrinsic threshold. For concreteness, if we assume that a value of  $F = 50$  is presented to all the alpha motor neurons. Only the neurons with thresholds from 1 to 50 are activated. Assuming, in our ideal world, all the alpha motor neurons have the same innervation ratio, activation of the 50 neurons probably produces 50 times the force generated by the activation of a single motor unit. However, the system rests precariously on the assumption that the thresholds of activations of the alpha motor neurons have a more or less uniform spread. The assumption that the thresholds of activation have a uniform spread is crucial for the size principle to work. Imagine the situation where all alpha motor neurons are equally excitable—all have a common threshold of, say, 50. Then any stimulation that is less than 50 fails to elicit any response from the alpha motor neuron system. But when the stimulation crosses the threshold of 50, all the alpha motor neurons respond. Both are extreme behaviors and are equally undesirable.

Another issue with the size principle stems from the fact that the muscle fibers generate force in bursts or twitches. They do not produce a uniform or constant contraction, unless they are stimulated by an input of high firing rate that puts the fiber in a state of tetanus. Imagine a hypothetical bundle of fibers all twitching at the same frequency. There are two extreme possibilities. If all the fibers twitch in perfect synchrony, like the oarsmen in the traditional *vallam kali* boat races of Kerala, the rhythmic forces of the individual fibers add up producing a huge fluctuating force at the level of the whole muscle (Fig. 9.10).

When you clench your fist to produce a stable force in one of the forearm muscles like the flexor digitorum, the fibers within could be twitching in complex rhythms—the total force generated, however, is felt to be stable and constant. To generate



**Fig. 9.10** A schematic illustrating the highly fluctuating net force generated with three fibers twitching (almost) synchronously



**Fig. 9.11** A schematic illustrating the generation of a stable net force with only a small ripple when the three fibers twitch asynchronously

a stable net force, though the individual fibers twitch rhythmically, the rhythms of the individual fibers must be out of sync among each other (Fig. 9.11).

This issue is lucidly expressed on page 559 of the classic 1991 neuroscience textbook by Eric Kandel, James Schwartz and Thomas Jessel as follows: “Although these (neural) firing rates produced unfused tetanic contractions, movements are smoothly executed because the different motor units are activated asynchronously.” What is the neural control mechanism that assures such desynchronized activation of muscle fibers? The size principle only addresses the issues of magnitude, not of the *rhythm*. The issue of desynchronized activity of motor units or muscle fibers was not paid adequate attention in muscle physiology literature, a lacuna that was pointed by Basmajian in his authoritative work on muscles: “The phenomenon of

motor unit synchronization has not been analyzed and documented as fully as other motor neuron properties...no detailed description of the behavior of synchronization as a function of measurable parameters, such as force and time, has been given."

In collaboration with P. S. Prashanth, the present author had earlier proposed the role of special spinal circuitry that can actively desynchronize motor unit activity. The motor control system in the spinal cord has several other classes of neurons in addition to the alpha motor neurons. For example, there are Renshaw interneurons to which the alpha motor neurons project; the Renshaw cells in turn supply inhibitory feedback to the alpha motor neurons. Thus, the alpha motor neurons inhibit each other via the Renshaw cells, a system generally referred to as the Renshaw inhibitory system. This inhibitory system is thought to keep the firing rates of the alpha motor neurons under control. But the role of the Renshaw inhibitory system is not just to suppress the firing rate of the motor neurons. With the help of a computational model, we have shown earlier that the inhibitory lateral interactions among the spinal alpha motor neurons are capable of desynchronizing their twitching rhythms. To support this claim, it has been shown that action potential recorded from distant parts of a contracting muscle are by and large desynchronized. On the contrary, in case of motor neuron disorders like poliomyelitis that affects the alpha motor neurons, large synchronized waves are seen in the Electromyogram (EMG), a system that measures the electrical activity of the muscle.

Above we have seen how the spinal neuron circuitry solves an important problem involved in the control of motor units and muscle units. But the spinal motor circuitry has several other important neurons and has more complex dynamics than what was described in this section. The next section presents the cellular elements of spinal motor circuitry and describes how it orchestrates the complex patterns of muscle activations that make our lives go.

## Spinal Circuits

Up to this point in the chapter, we persisted with an oversimplified description of the interaction with the spinal cord and the muscle. The interaction is described to be completely unidirectional: the neural master enslaving a hapless muscle. But the reality is quite different. The interaction between the nervous system and the muscle, as is the case with every other bodily system, is richly bidirectional. The muscles too send sensory signals about their state back to the neurons of the spinal cord. Without such feedback it would be impossible to control the muscles and produce reliable movements, as can be gathered from simple quotidian experiences.

Consider the task of holding a coffee cup, with your shoulder vertical, aligned with your torso, and the forearm extended in your front. The muscles controlling your fingers, wrist, elbow, and to some extent the shoulder, must produce precise levels of forces to counter the weight of the cup in your hand. To hold the cup level, the muscles of your hand and arm must contribute only a specific force—not more,

not less. For this to happen, the muscles must be able to communicate their forces (or tensions) to the spinal cord and constantly receive force corrections.

There are indeed special sensory organs, located not exactly in the muscle, that can transduce muscle tension and communicate it to the spinal cord. The muscles are firmly attached to the bones, not directly, but via intermediate structures called the tendons. Unlike muscle fibers, which are softer, a tendon consists of a tough band of fibrous connective tissue that connects the muscle to the bone. The fibers of tendons are made of collagen. The substance of the tendon is similar to the muscle in terms of its fibrous nature, and similar to the bone in toughness. For this reason, tendons are well fitted to bridge two very different types of tissue—muscle and bone.

At the junction of the muscle and the tendon, there are sensory receptors known as Golgi tendon organs. These are slender, encapsulated structures about 1 mm long and 0.1 mm in diameter. The tendon organs are innervated by free endings of nerves. When the muscle is stretched, the collagen fibers in the tendon straighten out thereby compressing the nerve endings, causing them to fire. Although tendons can undergo significant length change, particularly under high forces, the Golgi tendon organs are more sensitive to muscle tension, rather than the change in length. The electrical signals generated by the free nerve endings are carried by a class of fibers, known as the Ib axons, to the spinal cord.

There is another important variable of the muscle state that the nerve fibers communicate to the spinal cord. Imagine yourself responding to the question of the child who asks “How long is a foot?” You stretch out both of your hands, your palms facing each other, trying to mark an imagined foot in the air. For the hands to maintain a fixed mutual distance, the joints of the two hands (wrists, elbows, and shoulders) must be in a fixed configuration, and the muscles of the two hands must contract at fixed lengths. Therefore, to make the above performance, the muscles must communicate to the spinal cord, information about their current lengths.

Muscle length is transduced by special receptors called muscle spindles located in the fleshy core of the muscles. These are encapsulated spindle-like or fusiform receptors, each containing three components: (1) a set of specialized intrafusal fibers, (2) sensory nerve endings that originate from the intrafusal fibers, (3) motor nerve endings that innervate the intrafusal fibers. These intrafusal fibers run in parallel with the extrafusal fibers, what we have been referring all along glibly as “muscle fibers.” When the muscle stretches, and the (extrafusal) fibers elongate, the intrafusal fibers elongate too. When the extrafusal muscle fibers contract, the intrafusal fibers contract too, though the latter do not contribute much to the muscle force. This change in the length of the intrafusal fibers of muscle spindles, which reflects the change in the length of the whole muscle, is converted into electrical signals by the sensory nerve endings, the Ia axons, and communicated to the cord.

The motor nerve fibers that innervate the intrafusal fibers of the muscle spindles originate from the gamma motor neurons of the spinal cord. Whereas the gamma motor neurons innervate intrafusal fibers, the alpha motor neurons innervate our familiar (extrafusal) muscle fibers. Activation of gamma motor neurons causes shortening of parts of the intrafusal fibers. In such a condition, any shortening of the muscle

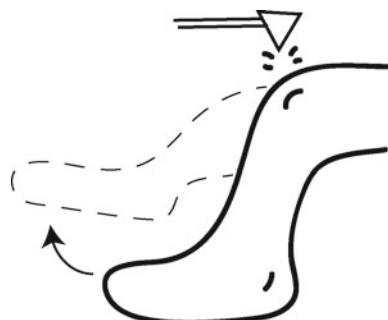
fibers causes increased firing of the Ia axons. Thus activation of gamma motor neurons increases the sensitivity of the intrafusal fibers in transducing muscle length.

There are two subtypes of the sensory Ia fibers. Let us simply call them dynamic and static Ia fibers, names that indicate their specific transductive properties. The dynamic fibers are sensitive not just to length but also to rate of change of length or velocity. The static fibers are more sensitive to muscle length. Interestingly, the sensitivities of these two fibers can be controlled by two corresponding gamma motor neurons. Increased activity of dynamic gamma motor neurons increases the sensitivity of the dynamic Ia fibers. Similarly, increased activity of static gamma motor neurons increases the sensitivity of the static Ia fibers.

Once we have laid out the motor commands from the nervous system to the muscle, and the sensory feedback from the muscle to the nervous system, we have the elements necessary to complete a *reflex arc*. A reflex is an automated or stereotyped movement elicited by a sensory stimulus given to the skin or the muscle. The sensory signal from the skin or a muscle propagates to the neurons in the spinal cord, immediately triggering a motor command back to the muscle, eliciting movement. A common example of a reflex is the patellar reflex, popularly referred to as a knee jerk reaction, clinically applied by the neurologist to test the health of lumbar segments, segments of the spinal cord that control targets in the abdomen and legs. To elicit this reflex, the clinician gently taps the patellar tendon under the knee cap (or the patella) with a rubber hammer; the subject reflexively extends the knee, making a kicking movement with the leg (Fig. 9.12).

The patellar reflex is an example of a type of reflex known as the stretch reflex, in which the sudden stretch of a muscle causes the muscle to contract in response. It was initially thought that this contraction is an intrinsic property of the muscle. But Sir Charles Sherrington, a preeminent British neurologist who worked in the late eighteenth and early parts of the twentieth century, demonstrated that the reflex is caused by the interaction between the spinal cord and the muscle. He showed that the reflex can be abolished either by destroying the sensory feedback from the muscle, or the motor fibers carrying the motor commands to the muscle. Sherrington was one of the first to recognize the significance of sensory stimulus in producing movement

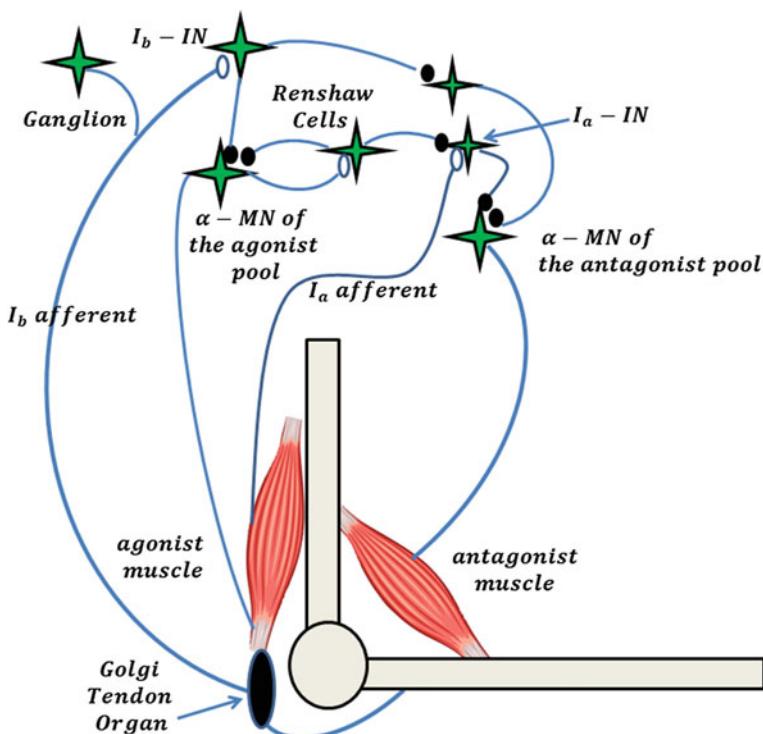
**Fig. 9.12** Knee reflex or patellar reflex



via reflexes. He envisioned that these basic reflexes are the building blocks of more complex movements.

Reflexes are mediated by neural loops that run between a muscle, or a set of muscles, and the spinal segments that contain neurons that control those muscles. We have noted above that sensory inputs from the muscle are carried by Ia- and Ib-afferent fibers (Fig. 9.13). These fibers are actually axons of neurons located in the dorsal root ganglion (DRG). These neurons are special in that both inputs and outputs are carried by axons. Sensory signals from the muscle propagating via the axons of these neurons, pass the cell bodies in the DRG, and proceed onwards to the spinal cord, synapsing, for example, on the alpha motor neurons in the ventral horn, thereby closing the reflex loop. A stretch of the muscle, therefore, sends signals that activate the alpha motor neurons, which in turn cause contraction of the stretched muscle—the series of events that underlie the stretch reflex.

More complex muscle–spinal neuronal loops exist. In another such loop, the Ia afferents from the muscle spindles synapse on the Ia interneurons in the spinal cord, which inhibit the alpha motor neurons of the antagonist muscles. Therefore, when



**Fig. 9.13** The spinal circuit consisting of  $\alpha$ -motor neurons ( $\alpha$ -MN), Renshaw cells, Ia, Ib interneurons (Ia-IN and Ib-IN), Ia afferents, Ib afferents, Golgi tendon organs, agonist and antagonist pair of muscles

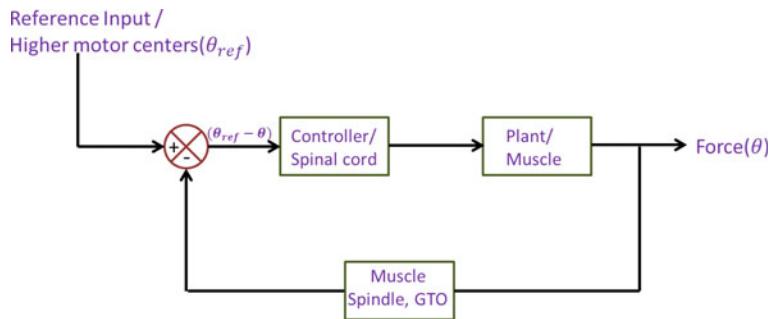
active alpha motor neurons cause contraction of the agonist muscle, signals related to the muscle length from the Ia afferents of the agonist muscle activate Ia interneurons, which inhibit the alpha motor neurons of the antagonist muscles. By the action of this loop, contraction of the agonist muscle, automatically cause relaxation of the antagonist muscle.

Another such loop involves the Ib-afferent fibers and the associated Ib interneurons in the spinal cord. Signals related to the muscle tension from the Golgi tendon organs of the agonist muscle, carried by the Ib afferents activate the Ib interneurons. These interneurons in turn inhibit the alpha motor neurons that innervate the antagonist muscle. By the action of this reflex loop, increased tension in the agonist muscle automatically reduces tension in the antagonist muscle.

The action of a spinal circuit on a single muscle, or on an entire section of the musculoskeletal system consisting of multiple muscles, joints, and bones, is best described by drawing analogies with a *servomechanism* from engineering. A servomechanism, or a *servo* for short, is a mechanism by which a physical quantity—like the angular position of a robotic arm, or the speed of a car—can be controlled so that it can follow a desired or a predetermined value. A robotic arm is controlled by a servo such that it reaches a desired angular position rapidly and accurately. A car run under cruise control is driven by a servo such that it moves at a constant speed, specified by the cruise control mechanism, despite the perturbations offered by the road and wind conditions.

Let us consider the components of a servo taking the example of the control of a robotic arm. The arm needs to be rotated from its current position to a desired position,  $\theta_{\text{ref}}$ . This turning is done by a motor located in its “elbow” joint. To this end, the motor must generate a torque that will rotate the arm exactly by the required angle. Application of a constant torque will not do, since under the action of a constant torque the arm will accelerate continuously gaining angular speed; it will not stop once it reaches the desired angular position. Therefore the torque generated must be positive for a little while, push the arm forward up to a point, and then turn to negative, braking the motion of the arm such that it comes to rest once it reaches the desired angular position. The problem is made more difficult, if we consider random perturbations to the arm. The control mechanism must correct the arm from the deviation caused by the perturbation and guide the arm stably to the desired angular position. Therefore, though the desired angle is a constant,  $\theta_{\text{ref}}$ , the actual signal to be given to the motor so that it achieves the control objective of taking the arm to that desired position, is quite complex. It is this that is accomplished by a servo.

Figure 9.14 shows a simplified block diagram of the parts of a servo. It basically consists of a loop, not very different from the loop of the reflex arc. The input to the entire loop is the desired quantity,  $\theta_{\text{ref}}$ . The output is the actual angle,  $\theta$ . The pathway from the input to the output is called the feedforward path; the reverse path from the output back to the input is called the feedback path. The objective of the servo loop is to ensure that  $\theta$  follows  $\theta_{\text{ref}}$  closely and quickly. The comparator depicted by “X” inside a circle, compares the desired ( $\theta_{\text{ref}}$ ) and actual ( $\theta$ ) angles and passes the difference (error =  $\theta_{\text{ref}} - \theta$ ) to the block named the “controller.” The controller processes this error signal and calculates the necessary torque, which is fed as input



**Fig. 9.14** Depicting muscle control as a servo loop

to the subsequent block labeled “plant,” which, in this case is a motor. The motor converts the input torque into the angle of the arm. The beauty of a servo is that though the reference signal,  $\theta_{ref}$ , is simply a constant, the complex torque profile needed to perform the necessary control is generated by the control block. This feat was made possible by the error signal, using which the controller determines how the torque must be adjusted from instant to instant. Since the servo achieves control by trying to minimize the error obtained by comparing the feedback signal with the desired signal, engineers call this form of control feedback error control.

Let us apply the feedback control framework to model one of the simplest functions of spinal motor circuitry. Figure 9.14 shows the stretch reflex depicted as a feedback control mechanism in which an alpha motor neuron (the “controller”) drives a muscle (the “plant”). The length of the muscle is fed via a muscle spindle (the “feedback module”) back to the alpha motor neuron. There is nothing explicit like a “comparator” in which the negative of the feedback signal is added to the reference input. The external input to the alpha motor neuron is the input to the spinal motor neurons from the higher motor centers in the cortex and the midbrain. Although there is no direct “comparison” of the feedback signal and the higher motor input, note that the feedback is a negative feedback since the stretched muscle triggers further activation of alpha motor neuron, thereby causing contraction. The feedback involved here is a negative feedback since increased muscle length goes around the loop, setting in motion an influence that reduces muscle length. Another interesting feature that must be added to this picture is the role of the gamma motor neuron. The higher motor commands also influence the gamma motor neurons, which affect the feedback branch (muscle spindles), varying the feedback gain. Although the interpretation of stretch reflex in terms of a feedback control loop brings a certain level of clarity into the operation of the reflex, it must not be taken too literally. The aim of this loop is not to maintain the muscle length such that it equals a length set point specified by the higher motor centers. The stretch reflex simply resists sharp changes in muscle length. The force with which a muscle resists being stretched is called the *muscle tone*, which refers to a certain tautness or firmness of the muscle.

More elaborate spino-muscular interactions can also be depicted using the general framework of servomechanism. In Fig. 9.14, the controller is the spinal circuit that consists of four types of spinal neurons—alpha and gamma motor neurons, Ia and Ib interneurons; the plant is the muscle itself; the feedback consists of the sensory information carried by Ia and Ib afferents. The input is the motor command conveyed by the descending pathways. In this case too, the servo interpretation cannot be taken too literally. There is no “error” signal that results from a comparison of the higher command and the feedback signal. The exact nature of the higher motor command is a subject of intense debate and cannot be simply interpreted in terms of a desired joint angle, muscle length, muscle tension, etc.

Perhaps the best way to use depictions like those of Fig. 9.14 is to convert them into mathematical models and analyze the flow of signals. Such models can be used to understand the relation between the higher motor commands and the resulting movement seen in the muscle length or joint space. Mathematical and computational models can take us to places where verbal descriptions, however articulate, will fail to do. One of the earliest models that depict how the spinal neurons described in this section control the agonist and antagonist pair of muscles of a single joint was proposed by Daniel Bullock and Stephen Grossberg. This model dubbed Factorization-of-LEngh-and-TEnsion (FLETE) model explains how the spinal circuit achieves accurate movement of a joint at variable force levels, at variable speeds. It essentially shows how force can be factored out of the joint control, while focusing only on the joint position.

Even elaborate models like the FLETE model do not begin to unravel the immense complexity of spinal motor circuitry. The spinal circuitry that we discussed so far in this section confines itself to a single spinal segment. The interactions among the spinal motor neurons and interneurons all occur within a single spinal level horizontally. But there are interactions that stretch across a large number of spinal levels. In addition, we have been describing spinal circuits playing a nearly passive role, waiting for the higher motor commands to breathe life in them, thereby animating the musculoskeletal system. But spinal circuits are capable of generating complex movements of the limbs without the higher commands informing every detail of the movement. There are interactions among distant spinal segments that control, for example, the upper and lower extremities. Such interactions drive the complex locomotor rhythms in four-legged animals, the subject matter of the following section.

## Spinal Control of Locomotion

The previous section on reflexes does not do complete justice to reflexes. Visualize for example the unpleasant experience of stepping on a pin, inadvertently of course. You (or your spinal cord actually) would withdraw your affected foot in a haste. But if that is the only local, hasty reflex that your cord can muster, you would have a nasty fall, adding to the pain. You would shift your weight to the opposite leg which is still in contact with the ground; you would tighten the antigravity muscles, muscles

that keep you erect defying gravity—of the unaffected leg so that it can sustain the whole body weight single leggedly; you would shift your body weight slightly backwards from your foot forward position; you might also extend the hands to further adjust your balance. All these complex, rapid movements can be orchestrated for the most part by spinal circuits, as a first line of response to the accident, though the higher motor commands kick in after a little delay. Execution of such whole body level reflex obviously cannot be driven by any one spinal segment alone; it needs involvement of a good number of spinal segments. A spinal segment receives inputs from the periphery via the dorsal horn; sends out its commands via the ventral horn; communicates with other spinal segments; receives inputs from the brain stem and higher up from the motor cortex.

At a first glance, locomotor rhythms like walking, running, sprinting, galloping, trotting, do not seem anything like the involuntary, sudden, jerky movements of a reflex. But Charles Sherrington believed that locomotor rhythms are conducted by chains of reflexes. His insights were not ungrounded, but have their roots in some elegant experiments he performed on the so-called *spinal animals*. Sherrington perfected a method by which the brain can be separated from the spinal cord by transecting the cord. In these spinal animals, the motor commands from the brain cannot reach the spinal cord and therefore the locomotor apparatus. Yet these animals showed stepping reflexes, rhythmic movements of the legs, when placed on a moving belt. The animal has to be supported, of course, since it cannot balance itself effectively without the involvement of the vestibular circuits located in the brain. Experiments with the spinal animals were certainly more elegant than earlier experiments with the same objective: dogs from which whole cerebral hemispheres were removed were able to exhibit walking behavior.

Soon after the transection of the cord in quadrupeds the animals become paralytic in the hind legs. But after a few weeks the hind legs recover stepping movements which can be sustained by training on a treadmill or general stimulation of the skin. Since external stimulation—contact with the tread mill or skin stimulation—seems to be necessary to induce stepping movements or locomotor rhythms, it appears that locomotion is indeed a reflex phenomenon. A reflex is essentially an automatic response to sensory stimulus. However, even when sensory signals are blocked in the so-called deafferentiated animals, locomotor patterns were observed. However, locomotor patterns generated in such extreme cases were much simpler than stepping patterns of normal animals. These results clearly demonstrated that locomotor rhythms are intrinsically and actively generated by the circuits of the spinal cord, though they can come under the influence of sensory stimuli. In that sense, locomotor rhythms are not composed of reflexes. There is a need for an alternative paradigm to explain the origins of locomotor rhythms.

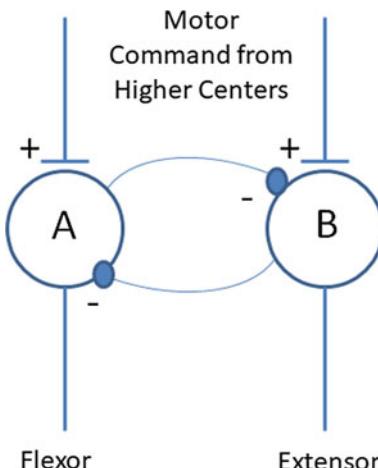
An alternative proposal came from Graham Brown who worked in the early part of the twentieth century. Brown observed that even the dorsal roots, the spinal regions that carry sensory inputs into the cord, are lesioned, alternating muscle contractions continue once they are triggered. Since the locomotor rhythms exist even in the absence of sensory inputs Brown argued that they are not reflexes. The rhythms are generated intrinsically by the spinal circuits, though it was not clear initially

how exactly the rhythms are generated. These hypothetical rhythm generating spinal circuits were dubbed the Central Pattern Generators (CPGs).

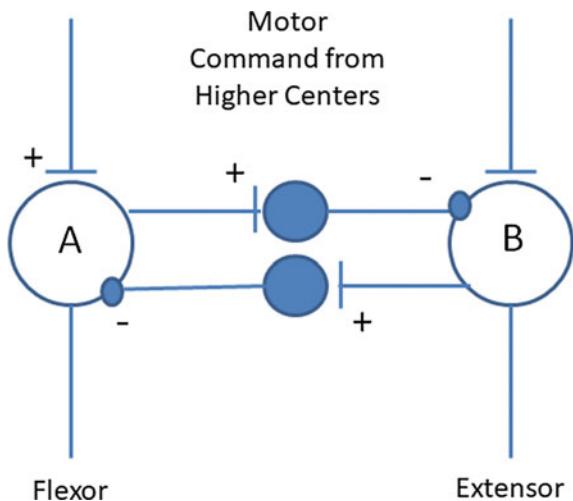
Brown proposed a theory of how the spinal circuits could be generating CPGs. Note that in the previous section, we only described spinal circuits controlling muscles on one side of the body. Brown considered spinal circuits controlling both sides of the body, which was obviously essential to account for locomotion in which there is coordination of limb movements on both sides of the body. He hypothesized that the spinal circuits on the two sides are coupled and are turned on and off alternatingly. At the heart of Brown's hypothetical circuit lie two mutually inhibiting neurons (Fig. 9.15). Imagine a neuron that is spontaneously active until it is inhibited. A neuron of that sort, A, is coupled to another neuron of the same kind, B, through inhibitory connections. Another required property of the neurons A and B is that they cannot sustain prolonged firing due to fatigue. Now let us consider a scenario in which A begins to fire first. B remains silent as long as A fires due to inhibition from A. But soon A stops firing due to fatigue. Now B springs back to life and fires. This time around B gets fatigued and stops firing after a while. The cycle continues. In Brown's spinal world, the motor neurons that control extensors (flexors) of one side inhibit motor neurons that control flexors (extensors) on the opposite side. This arrangement, which Brown calls *half-center organization*, generates a rhythmic motion in which the flexors (extensors) on one side move in sync with the extensors (flexors) on the other side, thereby generating a naturally observed locomotor rhythm.

In the '60s, Lundberg and coworkers in Sweden proposed a more physiologically grounded circuit model of Brown's theory. In their circuit, the motor neurons on the two sides of the cord do not directly inhibit each other; the inhibition is mediated by interneurons that play the role of half-centers (Fig. 9.16). In a more complex version of the circuit, Lundberg and coworkers added Ia interneurons and Ia afferents also to the original circuit.

**Fig. 9.15** Brown's spinal circuit with half-center organization



**Fig. 9.16** An extension of Brown's circuit by Lundberg and colleagues



There are other ways that a neural circuit can implement a CPG; a pair of mutually inhibiting neurons is only one of them. Some CPGs contain neurons that show intrinsic bursting behavior. Such neurons shoot a rapid train of action potentials in a burst and remain silent for a while. Single neurons generate such firing patterns; it is not an outcome of a network of neurons. Such endogenous bursters are ideal for driving autonomic rhythms such as the respiratory rhythms. In some cases, bursters are found to drive locomotor rhythms also. Some neurons generate an unusual pattern of membrane potential known as the *plateau potential*. Membrane potential in such neurons remains in a depolarized state for extended durations. The presence of such neurons in a neural circuit can also drive rhythmic activity at circuit level. Another neural mechanism that drives oscillatory activity is known as post-inhibitory rebound. Some neurons suddenly fire with active firing after the cessation of a spell of inhibition. When two such neurons are coupled by mutually inhibitory connections, the pair becomes capable of oscillations. Finally, administration of pharmacological agents also can trigger CPG-related activity in spinal neural circuits. For example, application of L-dopa, a substance used for synthesis of neuromodulator dopamine, can initiate walking movements in a spinalized cat. A similar effect can be seen by application of agonists of acetylcholine, another important neuromodulator. Application of the chemical *N*-methyl D-aspartate (NMDA) can initiate swimming movements in lamprey. Another key neuromodulator serotonin was also found to influence oscillatory activity of spinal neurons.

Above we have listed out several intrinsic mechanisms that influence CPG activity of spinal circuits, only to highlight the endogenous nature of these rhythms. But it would be a mistake to imagine that CPG rhythms are absolute rhythms, uncontrolled by external stimuli. If we invoke our personal experience with walking, we note that we constantly adjust our stepping pattern based on sensory feedback from the world, for example, the debris and the potholes on the road ahead. In a laboratory setting,

it may be observed that the speed of the treadmill drives the rate of stepping on a spinalized cat. Therefore, though the spinal CPGs generate an intrinsic rhythm, the rhythm is modulated by the sensory feedback from the peripheries.

Another major influence on the spinal CPG activity comes not from the world without but from above, from the higher motor centers in the brain stem and the cortex. In the '60s, Mark Shik, Fidor Severin, and Grigori Orlovsrky found that electrical stimulation of a midbrain center known as the Mesencephalic Locomotor Region (MLR) initiates stepping reflex in animals placed on a moving treadmill. The actual walking rhythm had nothing do with the pattern of stimulation but only its intensity. Walking was triggered when the stimulation intensity crossed a threshold level. Further increase in stimulation intensity produced increased walking speeds. In addition walking rhythms also changed with increasing speeds—from trotting at lower speeds to galloping at higher speeds. In trotting or normal walking, there is an out-of-phase relationship between left and right legs, which changes to an in-phase relationship in gallop.

Motor cortex too was found to have an influence on walking, though not on determining the basic walking rhythm. Lesions of motor cortex were found to impair walking tasks that involve a high level of visuomotor coordination like, for example, walking on a path with a lot of obstacles. Motor cortical neurons were found to be particularly active when animals were involved in such skilled walking situations.

Lesions of cerebellum also can cause impairment in walking movements. Motor impairments caused by cerebellar damage are collectively referred to as *ataxias*. Walking impairment in cerebellar damage patients takes the form of abnormal coupling between different limbs, or poor coordination between different joints. Like the spinal cord, cerebellum also receives sensory information from the muscles and sends out motor commands back to the muscles. But the key difference in case of cerebellum is that the sensory motor loop, if it can be so called, is a long one connecting several key higher motor areas. Therefore damage to cerebellum causes deep changes in walking rhythms, particularly in walking that involves visuomotor coordination. Damage to cerebellum also affects the sense of balance which is crucial for walking, particularly in bipeds.

In this section, we have seen a new face of spinal circuits. Far from responding to the shocks of the world by jerky, flailing movements, they are capable of generating complex locomotor rhythms that enable creatures to walk, run, swim, and fly. We have also seen how both sensory feedback and top-down influence from higher motor centers influence the rhythms generated in the spinal cord. In the following section, we will consider the contributions of higher motor centers to more general forms of movement.

## Motor Cortex and Willed Action

Control of movement by the higher motor cortical areas compared to the control exerted by the spinal cord is different in two fundamental ways. We have described

the movements driven by the spinal cord as reflexes, as though they are under helpless, uninhibited influence of the sensory stimuli. Under such circumstances, it is a bit odd to speak of motor “control” of the spinal cord, since control implies autonomy of some sort, an autonomy that is quite feeble at the spinal level since the motor output is under strong influence of sensory feedback. We have granted the spinal circuits some level of autonomy in the context of CPGs and locomotor movements, though, even in this case the output is strongly modulated by higher motor inputs and sensory stimuli.

Weaker autonomy is only natural at lower levels of the hierarchy in any large organization, since the freedom to decide usually rests with the highest levels of the management. The situation with the nervous system is no different. Since the motor cortex, or actually motor cortices since there exists a complex network of cortical areas engaged in motor control, along with the associated executive areas of the prefrontal cortex, represent the highest levels in the motor hierarchy, a forlorn dependence on the sensory input, a la the cord, would be unthinkable. The life experience of an individual in whom the highest motor decisions depend on a strong sensory determinism would be unenviable to say the least. Imagine the behavior of a person who responds to the stimuli of the world, completely devoid of inner censorship. If such a person sees a toothbrush, he would immediately pick it and start brushing himself; if he/she sees an apple, without a moment’s thought, he/she would grab it and take a bite; if he/she sees a bed, he/she would lay himself on it and make an untimely effort to enter states of sleep. Therefore, a very crucial element that must exist in our highest centers of motor control, an element that would rescue us from the damnation of an eternal zombiehood would be something that weakens the tight coupling between the sensory input and motor output. If there is a strong inevitable coupling between sensory input and motor output, a given sensory stimulus must be always, deterministically and helplessly, lead to a prespecified motor response. However, it is desirable that the sensory input is only *suggestive* of a certain motor output; there must be an internal mechanism, a *gate* of sorts, that determines if the sensory input goes through and precipitates the motor output. There must be an internal power of inhibition of a potential motor response. Or better, there must be a selectional mechanism that can choose from among several candidate motor responses in a given circumstance. Such gating is offered by the motor and prefrontal cortices in collaboration with other subcortical systems like, for example, the basal ganglia.

The ability to inhibit our responses, where the world demands it, might afford our behavior a greater social appropriateness and acceptability. The ability to choose from a set of possible actions perhaps elevates us from a robotic to a more human status. That makes us perhaps better than the robot but not yet fully human. Our actions are not always prompted by what we see around us here and now. Human spirit is defined by the uncanny ability to pursue far off dreams, dreams whose realization may sometimes demand long years of toil, incurring immense investments of energy and money, dreams that assume a living shape only in the mind of the dreamer, and not suggested, even in the faintest sense, by anything in the immediate sensory world. We like to work towards goals. We plan, strategize, and implement those plans by

performing actions whose outcomes unfold over long periods of time. A high schooler preparing for a major entrance exam subjects himself/herself, most often willingly, to one or two years of rigorous coaching. Any organism endowed with superior motor control must be capable of *goal-oriented behaviour*, a behavior that is again made possible by the motor cortex and the executive areas of the prefrontal cortex.

Let us begin our discussion of motor cortical areas with the primary motor cortex, since it is the simplest, though often made dangerously oversimplified. We briefly described the seminal work of Wilder Penfield on the motor maps in Chap. 1. Penfield was the first to popularize the presence of putative maps in the motor cortex, perhaps by presenting a dramatically simpler and therefore conceptually appealing picture of the nature of the motor cortex. But the earliest work on the motor cortex dates at least 70 years before Penfield's experiments.

In 1870 German neurologist Eduard Hitzig and anatomist Gustav Fritsch showed that electrical stimulation of certain parts of dog's cortex produced movements on the opposite side of the body. A few years later David Ferrier performed the first mapping experiments on monkey's motor cortex and found an approximate map of the body in the motor cortex. In this map, areas that controlled the feet are at the top and those that controlled the face are at the bottom. In the early twentieth century, the Vogt couple performed more detailed electrical mapping studies of the motor cortex.

In 1937, Penfield published some of his first historic studies of the motor maps. Penfield performed his studies as a part of surgical intervention for epilepsy that involved a search for epileptic focus using electrical stimulation. From his electrical mapping studies of the motor cortex, Penfield concluded, confirming some of the earlier studies, that there is a detailed upside down map of the contralateral body in the motor cortex. He noted that particularly large areas of the cortical real estate were allotted to the control of the hands and face, while larger parts of the body like the trunk have relatively smaller cortical allocation. He depicted these findings by drawing detailed two dimensional maps that resembled like an out of scale portrait of a human being. He called this being a homunculus (Latin for "little man") a symbolic creature that represents the action of the motor cortex. The motor cortex, thus anthropomorphised, caught the imagination of the students of neuroscience world over.

The motor map depicted by Penfield may be thought of as a convenient cortical keyboard, wherein each body part has a precise corresponding site from where it can be controlled. Thus we have two "motor keyboards," one in each hemisphere, controlling the muscles on the opposite side of the body. However, such a picture turned to be quite far from the reality. Studies performed by O. Foerster around the same time when Penfield published his first studies present a more complex account of the functional organization of the motor cortex.

In an article published in 1936, Foerster expounds the studies of British neurologist Hughlings Jackson on the motor cortex and highlights some of the key features of the functional organization of the motor cortex. Foerster's depiction of the motor cortex deviates considerably from that of Penfield and comes closer to contemporary descriptions.

One of the first organizing principle of the motor cortex, concerns with the relative cortical area allotted to a given motor function. In this respect, Foerster quotes Jackson: "The quantity of cortical grey matter varies not so much with the size of the muscles of a part of the body, as with the number of movements of that part. Thus the small muscles of the fingers will be represented by much more grey matter in the cortex than will be the voluminous muscles of the upper arm, because the former serve in more numerous, different, and in more specialized movements. Greater differentiation of function implies larger representation in the brain."

Contradicting popular accounts of the motor cortical organization, which assume that the motor cortex controls body parts on the contralateral side, Foerster states, again giving due credit to Jackson: "The second point is the bilateral cortical representation of different parts of the body...[muscles have] bihemispherical representation, that is to say, those of the two sides when acting together are represented in each of the two hemispheres, but also that every other muscle group is represented not only in the contralateral hemisphere, but to a certain degree in the ipsilateral hemisphere also." Thus, though the motor cortex on one side primarily controls the contralateral body, it also has an effect on the ipsilateral side.

The third organizing principle is that of overlapping foci. Penfield's motor maps present an impression that there are distinct points in the motor cortex that have distinct and exclusive effect on a certain body part. Foerster corrects this picture as follows: "...a single part of the body, let us say the thumb, is represented *preponderantly* in one part of the cortex, but it is represented in other parts of the precentral convolution as well, although in a different degree and in different combinations with other parts of the body." This one statement profoundly shatters the idea of a motor "map," which implies a point-to-point correspondence between the cortical surface and muscles of the body. Therefore, cortical sites where stimulation produces movement in a certain body part are not localized but distributed.

Just as the idea of a localized focus in motor cortical space warrants a correction, the idea of a brief stimulus, localized in time, producing a given uniquely corresponding movement is also a fiction and stands corrected. For example, Foerster states that repeated stimulation at the same cortical site, with a stimulus of the same intensity produces varied movement outcomes. "A given spot of the thumb-focus is stimulated repeatedly at intervals of one second by galvanic threshold stimuli. The first, second, third, and fourth stimulations result in a movement of the thumb. On the fifth stimulation this movement is fading. The sixth stimulus produces not a movement of the thumb, but a movement of the index finger. The same result is obtained by the seventh stimulus. On the eighth stimulation all fingers except the thumb move. On the ninth stimulation the finger movements are hardly visible. On the tenth and eleventh stimulations the hand moves, and on the twelfth stimulation the primary effect, the movement of the thumb, reappears and is obtained by subsequent stimulations." Such staggering variability in the effects of stimulation applied at a fixed locus completely flies in the face of map-like easy descriptions of the functional organization of the motor cortex.

More detailed stimulation studies reported in the '90s have confirmed the more sophisticated picture of the motor cortical organization described by Jackson, Foer-

ster, and others long ago. A given muscle can be activated by stimulating multiple sites in the motor cortex. Similarly stimulation at most sites activates multiple muscles, a functional observation that has been corroborated by careful track-tracing studies that show that axons from cortical neurons often terminate on widely distributed spinal motor neurons. Although modern observations force us to discard the simple keyboard like organization of the motor cortex, another interesting organizational principle that segregates proximal from distal muscles has been observed. Whereas cortical sites that activate distal muscles are concentrated in the center of a wider area, sites that activate proximal muscles are mostly distributed around that center on the peripheral areas.

The above accounts of functional organization of the motor cortex seek to link activity of single neurons to activity of single muscles. But when we move, our aim typically is not to produce pointed activations in specific muscles, of whose anatomical identities the uninitiated could be mostly ignorant. When we wave a hand, our goal would be to transport the hand from point A to point B. Are these movements coded in the motor cortex? If the answer is affirmative, in what format? This important question was taken up for study by Apostolos Georgopoulos and colleagues. In one of the first experiments of what turned to be an influential line of work, the team trained rhesus monkeys to make reaching movements from a center position to eight target positions in 3D space. Spiking activity was recorded simultaneously from a few hundred neurons as the animals moved their hands. Different neurons were active at different levels in different movements. On careful analysis, the team discovered that the activity of the entire *population* of neurons coded every movement direction. Each neuron represented a fixed direction in space; the activity level of that neuron represented the amplitude of an imaginary vector in that direction. When the vectors of all the neurons are added, the resultant vector accurately represented the actual direction in which the hand moved. The idea that populations of neurons code for a piece of information in the brain marks a significant step forward in our understanding of information coding strategies of the brain. In Chap. 7 on the visual system for example we noted how single neurons responded to oriented bars. The classical experiments by Hubel and Wiesel describe brain's representations of external stimuli in terms of single neuron responses. However, such an account gives the misleading impression that the firing of that single neuron exclusively represents the fact that a bar of a given orientation is presented to the brain. Actually there must be a large number of neurons in the neighborhood of that single neuron, all of which respond at various firing levels, to the same oriented bar. The identity of the bar stimulus is coded in the entire population of visual cortical neurons and not just a single neuron. Subsequent research by Georgopoulos and associates had revealed similar population codes in several other cortical and subcortical areas.

A long-standing question regarding the functional organization of the motor system concerns with the representation available at the motor cortical level in contrast to those at the spinal cord level. We have seen earlier that the alpha motor neurons of the spinal cord determine the force of contraction of muscles innervated by them. We have just noted that the motor cortical neurons code for movement direction. Based on such observations there have been attempts to arrive at facile and trenchant

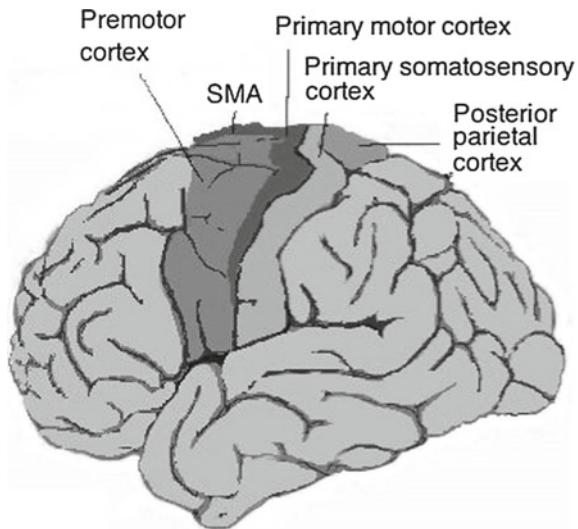
segregation of functions at the cortical and spinal levels. However, efforts to create such easy classifications have been thwarted by experiments that studied the effect of applied forces on motor cortical neurons. In one such study performed by E. Evarts with monkeys, the directional tuning of neurons, observed in the experiments of Georgopoulos and others, has been found to be modulated by external forces acting on the monkey's arm to either aid or oppose the movement. Forces that opposed the arm's movement resulted in increased firing rate whereas forces that pulled the arm in its tuned direction led to a reduction in firing rate. These studies showed that the representations generated in the motor cortex are complex combinations of kinematics and dynamics and not limited to kinematics alone.

The functional organization of the motor cortex is not only quite far from being strictly topographic, it is also quite dynamic, influenced by changing sensory stimulation, and not frozen forever. In one of the early experiments of its kind, John Donahue and colleagues have sectioned the facial motor nerve in adult rats and studied its effect on motor cortical organization. In rats, motor fibers of the facial nerve control, among other things, movements of whiskers. There is an entire area of the motor cortex that is dedicated to control of whiskers. Sectioning of the motor fibers that carry the commands from this motor area seems to have put this area temporarily out of employment. When the motor cortex is electrically mapped, a few weeks after the sectioning of the nerve, it was found that areas that originally controlled the forelimbs and eyelids have expanded, encroaching into the area that controlled the whisker movements. In some cases, the cortical changes were observed even 7 days after the nerve was sectioned. In Chap. 6 on brain maps, we have encountered several such examples of map reorganization that follow the principle of "use it or lose it."

The primary motor cortex is located next to the posterior boundary of the frontal lobe, running "north-south" parallel to the central sulcus, the prominent fissure that divides the brain approximately into anterior and posterior halves (Fig. 9.17). As we move further inwards into the frontal lobe, we encounter other, "higher" motor areas that project to the primary motor cortex. Two such higher motor areas have been identified: premotor area and supplementary motor area. The premotor area is located anterior to the primary motor area, and is further divided into dorsal ("upper") premotor and ventral ("lower") premotor areas. The supplementary motor area is located above the primary motor area, extending partly over the end of the hemisphere into the medial cortex where the cortical areas of the two hemispheres come into contact.

Earliest studies on the premotor cortex, that identify it as an area distinct from the primary motor area, date back to the early decades of the last century. However, these studies soon faced opposition from Wilder Penfield and others who denied the existence of a separate premotor cortex and believed that it is an extension of the primary motor cortex, part of a single motor map. However, the situation saw a turnaround in the '80s when efforts were made to clearly delineate the premotor and supplementary motor areas. One of the first attempts that sought to dissociate the functions of primary and supplementary motor areas was performed by P. E. Roland and colleagues using functional imaging techniques that measured the blood volume changes associated with neural activity. In this study, human subjects were

**Fig. 9.17** The locations of the primary motor, premotor and supplementary motor cortices



asked to make three kinds of finger movements: simple, complex actual sequential movements, complex imagined sequential movements. Simple finger movements produced activation only in primary motor area. When complex actual movements were executed activity was found both in the primary motor and supplementary area. When complex sequential movements were only mentally rehearsed without actual outward execution, only the supplementary motor area was active. The study revealed the relevance of the supplementary motor area to imagined movements and the necessity of primary motor cortex for manifest movements.

A similar study by Jun Tanji and colleagues reveals the difference between the premotor area and the supplementary motor area. In this study, monkeys were trained to perform different variations of sequential finger movements. In one variation, the monkeys were trained to touch, in a specific sequence, three panels visually presented in front of them. In another variation, they were cued to make the same finger movements from memory. When the monkeys merely followed visually presented panels their premotor area was activated. When their movements were internally driven, drawing from their memory, their supplementary motor area was activated. In both cases the primary motor cortex showed equal activation.

## Moving Willfully

In the last few pages, we have described some of the key motor cortical areas. Although anatomically a strict hierarchy does not exist among them, since both premotor and supplementary motor cortex project to the spinal cord, just as the primary motor cortex does, functionally there is a certain level of hierarchy. Whereas

activation of the primary motor cortex produces local contraction of part of a muscle, activation of the other two higher motor cortical areas produces a more complex sequential movement. But the highest and the most important function of the motor cortex is not simply to produce movements, simple or complex. There is another function of the motor cortex that makes it special among all brain areas, and that something makes us, the homo sapiens, special (provided certain deep philosophical conundrums are resolved) as a species.

We have seen that spinal cord too produces movement but more as a reflex, a rapid response, the content of which is informed by the stimulus itself. It is simply a stimulus-response loop, which is easy to understand in principle, though the details of it may be, on occasions, adequately complex. The premotor cortex also is distinguished by the fact that it is responsive to sensory stimulus: motor commands from the premotor cortex are strongly determined by the sensory information streaming in from the posterior brain into the motor cortex. But the supplementary motor area is different. It is known to be particularly active in case of self-generated movements.

The expression “self-generated movement” like its kindred expression “voluntary movement,” is used quite normally in neurobiology and in neurology literature, as if everyone concerned (the writer, who used the expression in a book or an article, and the reader alike) knows exactly what they are talking about. All of us are quite familiar with the fact that we can originate movements. We can extend our hand, at a self-chosen moment, and grab that pen that we see in front of us sitting idly on the table. We can tap on the keys of the keyboard in front of us, in the most preposterous sequence that we can imagine, and produce the most impossible monstrosity of text ever produced. This primal ability to do things on our own, to reach out, touch, stand up, sit up, walk, and stop at will, to think our own private thoughts, to dream our own surreal dreams, to have a will of our own—this is what makes us what we are as human beings. Bereft of this primordial power we are stripped of our humanness, reduced to the state of a vegetable.

This basic power that we acknowledge as the power of will or volition in our common quotidian experience is referred to as the problem of the free will in philosophical circles. It is a problem because it is not totally clear, philosophically speaking, if it indeed exists. What I accept and swear by constantly as a conscious individual, what I consider to be the *alpha* of all my experience, this essential I-ness that forms the bedrock of all that I call myself, is sometimes considered, when subject to the exquisite methods of philosophical scrutiny, simply a delusion, an erroneous belief, a bad idea.

We can argue philosophically ad infinitum about the existence or non-existence of free will. But debating free will, thankfully, is not an exclusive prerogative of the philosopher and the metaphysicist. Someone like a physician, who has to deal with the hard and sometimes disgusting reality of matter and mucus, is often faced with the challenge of the free will. Free will is clinical problem, a challenge of clinical neuroscience, and therefore deserves the serious attention of the neuroscientist.

Perhaps one of the first thinkers who moved the free will debate—he used the expression “willed action”—from philosophy to psychology, was American philosopher and psychologist William James. James classified movements generated by

the brain into two types: “ideo-motor” and “willed” actions. Regarding ideo-motor actions, he says: “wherever movement follows unhesitatingly and immediately from the notion of it in the mind, we have ideo-motor action. We are then aware of nothing between the conception and the execution.” On the other hand, in case of willed actions, he says that there is “an additional conscious element in the shape of a fiat, mandate, or expressed consent.” James also pointed out a distinguishing feature of will: “Effort of attention is thus the essential phenomenon of will.” In other words, will and attention are two faces of the same coin.

The fact that willed action is disturbed in certain neurological and neuropsychiatric conditions brings the problem closer home, making the phenomenon of willed action amenable to the methods of neurobiology. One such instance is the Parkinson’s disease, a neurodegenerative disorder that was originally thought to be a motor disorder, but was subsequently found to affect all the four major domains of brain function—motor, cognitive, affective or emotional and autonomic (referring to the automatic neural control of the internal organs). Parkinson’s patients typically exhibit slowness of movements, a feature known as bradykinesia. They often also have trouble initiating movement, a condition in more extreme situations worsens into akinesia or total lack of willed movements. Parkinsonian akinesia has been described by Kinnier Wilson as a “paralysis of will.”

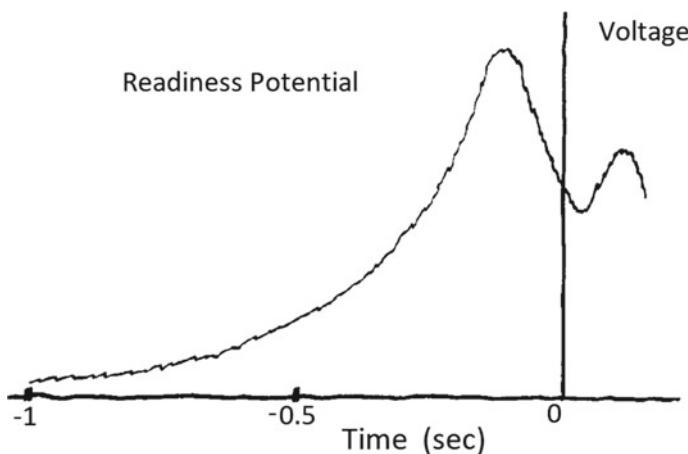
Some Parkinson’s patients exhibit an impaired gait characterized by short shuffling steps, a type of gait known as festination. They also walk with a precarious stoop that seems to make them more prone to falls. This impaired gait is sometimes punctuated by sudden halts, a phenomenon known as freezing of gait. The patient suddenly stops in his/her tracks as though he/she momentarily forgot how to take that next step. Apparently innocuous stimuli can trigger the freezing of gait. Like having to walk through a narrow corridor, or having to make a sudden turn and so on. Any cognitive distraction while they are walking can also trigger freezing. Once they get going, they continue to move, but the trouble lies with the initiation. There are strong reasons to believe that these problems of gait reflect a fundamental difficulty faced by these patients in translating will into action.

The problem assumes a more dramatic form in Parkinson’s patients suffering from the so-called hemi-parkinsonism. In these patients there is bradykinesia on the affected side, while the intact side exhibits normal willed movements. Another peculiar feature of this disease, dubbed “paradoxical kinesis,” is that while the patients have difficulty in initiating movements by their own will, they can sometimes exhibit significant movement like walking or running when challenged by a risky situation like fire or a similar hazard. These clinical observations reveal that, what is affected in Parkinson’s patients is not really the motor machinery that generates movements. The link that connects will to that machinery is broke. The problem can be described from the patients’ personal point of view also—they know what they want to do but cannot do it.

The above definition of willed action is quite abstract. It is not an operational or practical definition of willed action. Applying the empirical and objective standards of science, how does one classify a given performed act as willed or stimulus-driven? Does the brain discriminate between the two kinds of actions? Technically speaking,

what are the neural substrates of willed action? An affirmative answer to this question emerged from the studies of Hans Kornhuber and Luder Deecke in 1964. The scientist duo recorded Electroencephalogram (EEG) data as subjects performed self-initiated movements. The scientists were intrigued to find out that, in case of self-initiated movements, an electrical potential begins over the top of the head on the midline (close to what is called Cz electrode in the EEG system) nearly one full second before the onset of actual movement. It is as though the brain is making elaborate, time-consuming preparation for the self-initiated movement. Such an electric potential was not observed in stimulus-driven movements. This intriguing potential was given a tongue twisting German name called the Bereitschaftspotential (Fig. 9.18). The English-speaking world refers to it as the “readiness potential.” It is a signal that the brain is getting ready to produce movement. Interestingly, the readiness potential is weak when Parkinson’s patients try to move voluntarily, mirroring the familiar difficulty they face in initiating movement.

Computational “dipole analysis” that estimates what electrical sources in the brain produce what EEG components predicted that the readiness potential is generated by the Supplementary Motor Area bilaterally. Although the activity is produced initially bilaterally, after a little while (a few hundred milliseconds) the activity gets localized to the contralateral motor cortex, before the movement begins. We have already learnt that activity in the contralateral motor cortex drives the movements of the body parts. What causes that motor cortical activity is the readiness potential that is bilateral. That answers, though only partly, a key question that arises in case of self-initiated movement. In case of stimulus-driven movement, it is the external stimulus that produces a chain of events in the brain. But in case of self-initiated movement, where is prime mover in the brain? Who starts the neural fire that drives voluntary movement? The quick answer is the bilateral Supplementary Motor Areas.



**Fig. 9.18** The readiness potential or Bereitschaftspotential building up towards point of movement onset (0 s)

What is more intriguing about the willed-action signal, to borrow James' expression—is not the “where” of it but “how.” The source of the willed-action signal could have been some other brain region, like a prefrontal area for example, if not the Supplementary Motor Area. But what is so distinguishing about it is the slow build of activity that is characteristic of the readiness potential. Aaron Schurger, Jacobo Sitt, and Stanislas Dehaene have proposed a model to explain the generation of the readiness potential. They called their model the “stochastic accumulator” model. Basically the creators of this model have assumed that there is quantity that is growing by spontaneously accumulating into itself a constant quantity. For example let us begin with an initial value of  $x(0) = 3$  which accumulates into itself a constant quantity  $a = 2$ . The value  $x$  now grows as:  $x(1) = x(0)+a = 3+2 = 5$ ;  $x(2) = x(1)+a = 5+2 = 7$ ;  $x(3) = x(2) + a = 7 + 2 = 9$ , and so on. To this simple linear growth we need to add two more features. Add a random quantity at every step so that growth process shows random deviations from the straight line. The other feature a mechanism of decay: at every step allow  $x$  to lose a fixed fraction of itself. The actual movement onset occurs whenever this seemingly random variable reaches a fixed threshold or set point. The time at which the signal reaches the threshold varies from trial to trial due to the noise added to the growing signal. With this simple model Schurger and his colleagues were able to explain the exact form of the readiness potential fairly accurately. But the model is still a strictly numerical model and does not give insights into how different brain structures produce readiness potential.

The present author had proposed a model of the willed action that describes the possible anatomical underpinnings of the readiness potential. To understand the origins of the readiness potential one must first consider the so-called motor preparation which refers to the activities in the brain that occur as a part of the brain's preparation for the impending movement. The early activity seen in the bilateral Supplementary Motor Areas, for which the readiness potential is a readout, is only *one* preparatory process. Only that particular preparatory signal was picked up in the EEG because EEG electrodes are close to the surface and can pick up only cortical activity more prominently.<sup>1</sup> But electrophysiological studies that probed the deep brain structures for motor preparatory activity have found such activity in the structures of basal ganglia also—the striatum, subthalamic nucleus, and so on.

The present author had proposed according to which the readiness potential is generated by the interaction between the motor cortical areas and the basal ganglia. The essence of this theory is that there is a *willed action signal* that probably arises in the highest centers of cortical hierarchies in the prefrontal cortex. This putative signal is, by its very nature weak, and therefore needs special mechanisms for amplification, for it to be expressed into motor action. This amplification is done by the loop of interactions between the cortex and the basal ganglia. The initial weak willed-action signal circulates through the cortico-basal ganglia loop and gradually gets charged up. The build-up of signal that is described in the aforementioned accumulator model occurs by the dynamics of the cortico-basal ganglia loop. The noise that is added

---

<sup>1</sup>There are of course the high-density EEG systems that use hundreds of electrodes from which activity of deep brain structures can also be estimated using sophisticated computational models.

in the growing signal is proposed to arise from the special part of the basal ganglia known as the subthalamo-pallidal loop. This loop consists of two nuclei: the Subthalamic Nucleus, an excitatory nucleus, and the Globus Pallidus externa an inhibitory nucleus. A loop of excitatory and inhibitory nuclei of this kind is known to be capable of generating complex neural dynamics, which in the present case acts like a source of noise. The thresholding step in the accumulator model probably occurs in the thalamus. The output of the basal ganglia has to go through the thalamus before it returns to the motor cortex. In the accumulator model, when the growing signal reaches a threshold, it is as if a gate is opened and movement ensues. A similar gate to motor action is thought to be located somewhere at the meeting point of the basal ganglia output and the thalamus. This gate is shut tight in conditions like Parkinson's disease where there is a difficulty in initiating movement. Surgical treatments to relieve such a condition include lesioning of the output nucleus of the basal ganglia—Globus Pallidus interna—which is tantamount to breaking the gate open. A more detailed and rigorous account of this theory that can explain a wide variety of functions of the basal ganglia is possible. But it is beyond the scope of the present book.<sup>2</sup> In summary, we believe that the willed-action signal is a weak signal originating from the highest centers of motor and executive cortices. This signal needs amplification for it to get expressed and manifest as movement. This amplification occurs as the originally weak signal circulates through the cortico-basal ganglia loop until it reaches a threshold.

The above theory places the accumulator theory on a more concrete anatomical and dynamic foundation. It is also consistent with a wide variety of functions of the basal ganglia as described in the book by Chakravarthy and Moustafa (2018). But there are still open questions regarding the nature of the willed-action signal. One of the popular positions about free will in contemporary neuroscience is that there is no such thing. Free will is considered as an illusion. Seminal experiments by Benjamin Libet on the timing of the willed action with respect to the readiness potential lead us to the tantalizing conclusion that the readiness potential begins to build up even before the subject has the conscious feeling of the willed action. The experiment makes the startling suggestion that the brain initiates movement all by itself and only informs the subject, the proud owner of the brain, only by way of respect. Some of these issues are discussed in greater detail in the last chapter on Consciousness. Free will is a real phenomenon in both common daily experience and in clinical conditions of willed action and its impairments. But it is an illusion in certain philosophical positions. We do not have a satisfactory resolution of the mystery at this point. But just as an discussion of sensory processing leads in its advanced stages to the knotty issues of sensory awareness, any discussion of motor function in its advanced stages will have to necessarily lead on to the unresolved challenge of the free will. We have religiously led the discussion of motor function to its inevitable foggy boundaries. We await future developments in motor neuroscience to render those borders unambiguous and satisfactory.

---

<sup>2</sup>The reader may kindly refer to Chakravarthy (2013) and Chakravarthy and Moustafa (2018).

## References

- Basmajian, J. V. (1962). Muscles alive. Their functions revealed by electromyography. *Academic Medicine*, 37(8), 802.
- Biewener, A. A. (2011). Muscle function in avian flight: Achieving power and control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1570), 1496–1506. <https://doi.org/10.1098/rstb.2010.0353>.
- Brown, T. G. (1911). The intrinsic factors in the act of progression in the mammal. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 84(572), 308–319.
- Bullock, D., & Grossberg, S. (1988). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review*, 95(1), 49.
- Chakravarthy, V. S. (2013). Do Basal Ganglia amplify willed action by stochastic resonance? A model. *PLoS ONE*, 8(11), e75657P.
- Chakravarthy, V. S., & Moustafa, A. A. (2018). *Computational Neuroscience Models of the Basal Ganglia*, Springer Verlag, Singapore.
- Dick, J. P. R., Rothwell, J. C., Day, B. L., Cantello, R., Buruma, O., Gioux, M., et al. (1989). The Bereitschaftspotential is abnormal in Parkinson's disease. *Brain*, 112(1), 233–244.
- Foerster, O. (1936). The motor cortex in man in the light of Hughlings Jackson's doctrines. *Brain*, 59(2), 135–159.
- Gemmell, B. J., Costello, J. H., Colin, S. P., Stewart, C. J., Dabiri, J. O., Tafti, D., et al. (2013). Passive energy recapture in jellyfish contributes to propulsive advantage over other metazoans. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1306983110>.
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233(4771), 1416–1419.
- Jahanshahi, M. (1998). Willed action and its impairments. *Cognitive Neuropsychology*, 15(6–8), 483–533.
- James, W. (1890). *Principles of psychology*. London: MacMillan.
- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (1991). *Principles of neural science* (3rd ed., p. 559). New York: Elsevier.
- Kornhuber, H. H., & Deecke, L. (1964). Hirnpotentialanderungen beim Menschen vor und nach Willkürbewegungen dargestellt mit Magnetbandspeicherung und Rückwärtsanalyse. *Pflügers Archiv-European Journal of Physiology*, 281(1), 52.
- Latash, M. L. (2008). *Neurophysiological basis of movement*. Champaign: Human Kinetics.
- Loeb, G. L., & Chez, C. (2000). The motor unit and muscle action. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (Vol. 4, Chapter 34). New York: McGraw-Hill.
- Lundberg, A. (1981). Half-centres revisited. In *Advances in physiological sciences. Regulatory functions of the CNS: Motion and organization principles*. J. Szentagothai, M. Palkovits, & J. Hamori (Eds.), Vol. 1 (pp. 155–167). Budapest: Pergamon/Akademiai Kiado.
- Nussbaum, M., & Nussbaum, M. C. (1985). *Aristotle's De Motu Animalium: Text with translation, commentary, and interpretive essays*. Princeton: Princeton University Press.
- Pearson, K., & Gordon, J. (2000). Spinal reflexes. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (Vol. 4, Chapter 36). New York: McGraw-Hill.
- Penfield, W. (1961). Activation of the record of human experience: Summary of the Lister oration delivered at the Royal College of Surgeons of England\* on 27th April 1961. *Annals of the Royal College of Surgeons of England*, 29(2), 77.
- Prashanth, P. S., & Chakravarthy, V. S. (2007). An oscillator theory of motor unit recruitment in skeletal muscle. *Biological Cybernetics*, 97, 351–361.
- Ruppert, E. E., Fox, R. S., & Barnes, R. D. (2004). *Invertebrate zoology* (7th ed., p. 82). Pacific Grove: Brooks/Cole.

- Schurter, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of National Academy of Sciences*, 109(42).
- Shik, M. L., Severin, F. V., & Orlovsky, G. N. (1969). Control of walking and running by means of electrical stimulation of the mesencephalon. *Electroencephalography and Clinical Neurophysiology*, 26(5), 549.
- Smith, K. K., & Kier, W. M. (1989). Trunks, tongues, and tentacles: Moving with skeletons of muscle. *American Scientist*, 77(1), 28–35.
- Vogel, S. (2003). *Prime mover: A natural history of muscle*. New York: WW Norton & Company.

# Chapter 10

## Circuits of Emotion



*Emotion always has its roots in the unconscious and manifests itself in the body.*

—Irene Claremont de Castillejo.

A book on brain is incomplete without a discussion of what emotions are, where, if they can be localized, they are located in the brain, and how brain handles them. Emotions pose a peculiar problem to the neuroscientist. They are vague and elusive. They evade precise definition and rigorous characterization. There are other aspects of brain function, like the sensory-motor function, for example, that are a child's play to the neuroscientist compared to the challenge offered by emotions. Vision is doubtlessly a complex problem. How does the brain process form, static and dynamic, or color and other primitive properties? Or how does it identify a complex entity like the grandmother? Which parts of the brain participate, in what precise fashion, when I scan a crowded image for a familiar face? These are obviously difficult questions, because the domain of study is complex and involves immense detail; not because it is vague. Similar comments could be made on other types of sensory function—hearing, touch, smell, and taste. We can split sounds into frequencies, or categorize touch in terms of light, deep, or vibratory touch. We can describe the chemistry of smell and taste. Likewise, our motor system, the part of the brain that generates our movements, offers no difficulty in basic definition, quantification, and measurement of motion. The difficulty lies in the extraordinary detail involved in describing movements and the circuits that control them. But such is not the case with emotions. Where do we begin in our search for a science of emotions? Do we seek out brain areas of love and hate? Do we look for neurons whose firing rates code for precise intensities of the spectrum between pleasantness and ecstasy? Is there a hierarchy of emotions, with primary emotions represented in the lower layers, and more complex emotions in the higher layers? Such naïve line of questioning would have worked with lesser aspects of brain function. But, if our aim is to get a neurobiological grip on emotions, we must tread more carefully.

## Ancient Emotions

Although emotions themselves are as ancient as man and mind, a science of emotions, in the sense of modern Galilean science, is perhaps only over a century old. Earlier explorations into the world of emotions occurred in the domains of literature, poetry, philosophy, art, culture, and even religion.

There is no precise translation for the word emotion in ancient Indian philosophy. A large body of original Indian philosophical literature was written in Sanskrit. The Sanskrit terms that come closest to emotion are *bhava* (feeling) and *samvedana* (sensation/experience). But *bhava* has a manifold connotation and can be loosely translated as mood, outlook, perspective, or even attitude. Bhagavad Gita, an essential text in Indian spiritual literature, comments on emotions in quite negative terms when it refers to desire (*kama*), anger (*krodha*), greed (*lobha*), delusion (*moha*), pride (*mada*), and jealousy (*matsarya*) as the six inner enemies that must be identified and vanquished. It exhorts the individual to shun love (*raga*), or the attachment that it causes, as much as hatred (*dvesha*) since both are two sides of the same coin, and ultimately lead the individual to attachment to the object of love/hate, culminating in sorrow and bondage. The only love that is approved and admired is the love that is directed toward God, and such love is discussed and described at great length in various ancient Indian writings. The Narada Bhakti Sutras, a treatise on devotional love, speaks of nine stages of blossoming of love turned toward God. Patanjali yoga sutras, a treatise on systematic inner development, warns that love (*raga*) and hate (*dvesha*) are afflictions (*klesa*) of the mind, impediments to spiritual progress. Thus, due to its dominant preoccupation with a goal that is otherworldly, Indian philosophy does not seem to indulge in emotions but only talks of their transcendence and sublimation.

But Indian theory of aesthetics, the theory of rasas, seems to take a more considerate and inclusive view of emotions. The word *rasa* means “juice” literally, but is used in the sense of “essence,” the essential qualities and colors of experience. The theory of rasas, which first appears in Natyashastra, an ancient treatise on the science of dance and drama, speaks of eight primary rasas. These are love (*sringaram*), humor or mirth (*hasyam*), fury (*raudram*), compassion (*karunyam*), disgust (*bibhat-sam*), horror (*bhayana-kam*), valor (*viram*), and wonder (*adbhutam*). Each of these rasas or emotions is associated with a color and even a deity. For example, the color of love is light green (not pink!) with Vishnu, the god of preservation and sustenance, as its presiding deity. The color of terror is black, presided over by Kala, the god of death and destruction. To the list of eight rasas, a ninth—known as *shantam*, which stands for peace and tranquility—was added around the eleventh century. Two more—*vatsalyam* (love or fondness of a senior person toward a junior) and *bhakti* (devotion to God)—were added subsequently. The evolving list of emotions in Indian tradition shows that there is finality to the list.

Western philosophy seems to grant to emotions a more consistently respectful status. Plato, one of the great thinkers of ancient Greece, describes, in his Republic, that the human mind has three components—the reasoning, desiring, and emotional

parts. Plato's student Aristotle, with his penchant to pronounce upon things at length without any objective support, gave a long list of emotions: anger, mildness, love, enmity (hatred), fear, confidence, shame, shamelessness, benevolence, pity, indignation, envy, emulation, and contempt. Spinoza a philosopher of seventeenth century, with strong theological leanings, posits that emotions are caused by cognitions. Affects, the word that Spinoza used for emotions, like hate, anger, envy, etc., follow their own laws just as everything else in nature. He rejected the notion of free will, since the will, which is presumed to be free, has a hidden cause, which in turn has another cause, and so on ad infinitum.

## Emotions in Psychology

Thus, philosophical or aesthetic inroads into the subject of emotions were based on introspection, insight, and speculation and often lack an objective basis. Therefore, the number and classification of emotions had no finality or definiteness and varied with place, epoch, and cultural milieu. But then the need for an objective basis and a universal framework is a peculiar need of modern science and does not constrain art or philosophy. Even a preliminary attempt to find universal emotions must go beyond common cultural knowledge and anecdotal information arising out of immediate nativity, and warrants a comparative study of emotions in a range of world cultures. Keeping in line with the traditions of objectivity, attempts were made to classify emotions based on facial expressions, which can serve as sensitive markers of emotions. Based on a study of universal patterns in facial expressions, Sylvan Tomkins had arrived at a list of eight basic emotions—surprise, joy, interest, fear, rage, disgust, shame, and anguish. Although it is tempting to compare some of these emotions with the rasas of Indian aesthetics (rage = raudram; fear = bhayanakam, etc.), one can easily get carried away by such analogies. Since all cultures share the same neurobiology, it is not surprising that there are some emotions shared by all. But the difficulty arises if we seek a uniquely universal list of emotions. Based on analysis of universal facial expressions, Paul Ekman proposed the following six basic emotions: surprise, happiness, anger, fear, disgust, and sadness. The close resemblance to the typology of Sylvan Tomkins is easily noticed.

An interesting attempt to organize emotions hierarchically, not relying completely on facial expressions as the basis, was made by Robert Plutchik. In Plutchik's system, there are basic emotions and their combinations which generate "higher order" emotions. There are eight basic emotions arranged in the form of a circle (Fig. 10.1). Each of the basic emotions has a corresponding basic opposite emotion (joy—sadness, fear—anger, and so on). Angular distance of emotions on the circle is a measure of their similarity—nearby emotions are more similar. The basic emotions can be combined, a pair at a time, to produce mixed emotions called dyads. Blends of emotions that are at adjacent positions on the circle are called first-order dyads. The blend of joy and trust/acceptance corresponds to friendliness. Fear and surprise produce alarm. Combinations involving emotions with one other intervening emotion

**Fig. 10.1** Robert Plutchik's wheel of eight basic emotions



are called second-order dyads. For example, a combination of joy and fear results in guilt. There are also third-order dyads constructed by combining basic emotions with two spaces between them. Plutchik's system was able to accommodate a good number of complex and subtle emotions in an elaborate framework.

But for the fact that there is no objective, neurobiologically rooted, quantitative basis, Plutchik's system of emotions gives a considerable insight into interrelationships among emotions. It might serve as a guideline, a map for any future endeavor directed toward creation of a comprehensive neural theory of emotions. Its placement of emotions on a wheel, with opposite or complementary emotions located on opposite ends of the wheel is reminiscent of the "color wheel" used by artists to comprehend color. Colors too are organized in a simple circular map—the color wheel—and segregated into primary, secondary, and complementary colors. In fact, there are several such systems. In one such system, known as the red-green-blue (RGB) system, the primary colors are red, blue, and green. Their complementary colors are cyan, magenta, and yellow, respectively. To test this, stare at a red square for a little while (about 30 s) and then shift your gaze to a white background. You will see an afterimage of the red square, which will turn out to be a cyan square, hovering in your sight. Cyan is what you get when you subtract red from white, and is therefore complementary to red. It is tempting to compare this transformation of a color into its complement, to a similar conversion of emotions, to the manner in which love, when spurned, turns into its opposite, hate. These distracting speculations aside, color classification of the kind mentioned above is based on extensive study of human visual perception, a field known as visual psychophysics. In addition, these studies are also corroborated by the study of responses of the photoreceptors in the eye, and a whole range of neurons spread over the visual hierarchy from the retina to higher visual cortical areas. Perhaps, Plutchik's classification is actually fashioned on the lines of color theory. But then such correspondence is at the best

an analogy, an insightful metaphor, and nothing more. A theory of emotions that is rooted in the brain will necessarily have to be a very different beast.

How do we begin to construct a neural theory of emotions? How do we tether the tempests of emotion to the calm, motionless ground of the brain? To begin to answer these questions we must, first, notice a serious omission in the aforementioned approaches to emotion. Emotions are basically very different from thoughts. The contents of an emotional experience are very different from the contents of a sensory-motor experience. The products of cognition can perhaps be neatly segregated into bins, with convenient labels. Anyone who has struggled with the difficulty of figuring why they feel what they feel in certain emotional moments knows that emotions do not lend themselves to such easy analysis. A deep reason behind this difficulty is that emotions do not limit themselves to the brain and mind—they spill over into the body, and demand its implicit participation. When a mathematician is lost in thought in the tranquil solitude of a pleasant night, her brain is perhaps feverishly active but the body might remain calm, motionless, allowing the brain its full play. But when a young man struggling to utter his first words of endearment to the girl of his love, what he experiences is a tumultuous state of mind that is accompanied by the pounding of the heart, the sweating of palms, frozen and disobedient limbs, the flushing of the face, dilated pupils, and so on. It is as though the whole body is struggling to express those first feelings of fondness. Describing the outward signs of a devotee experiencing divine ecstasy, Vaishnava devotional literature mentions sudden perspiration, choking, tears, and horripilation. Thus, our cogitations are purely mental, cerebral. Our emotions, on the other hand, carry the brain and the body in a single sweep.

We now turn our attention to the nature of the bodily changes that accompany an emotional experience. Accelerated cardiac activity, perspiration, and dilation of the pupils are effects of a part of the nervous system known as the sympathetic nervous system. It coordinates what is described as a flight-or-fight response in the body. When an animal prepares to fight a predator and defend itself, its sympathetic systems try to muster all its somatic resources in a desperate attempt. Pupils are dilated so as to enable the animal to take in as much visual information as it can to aid its defense. Heart accelerates to meet the additional energy demands of the body engaged in fight. Perspiration in the skin increases so as to shed extra heat produced. Therefore, in addition to the cognitive information about the object of the emotion, be it love, hate, anger, or fear, the emotional experience involves a whole range of sympathetic effects in the body.

Therefore, we observe that emotional experience consists of two components: a cognitive registration of the object of emotion, the loved one, or a fearful predator, and so on, which serves as a stimulus for the emotion, and the bodily response. But where do we place the feeling that goes with emotion? Is the feeling of panic or love produced by the first contact with the stimulus, or does it develop as a result of the bodily response? In other words, is the feeling a result or a cause of the bodily response? This interesting chicken-and-egg question about the origins of emotional feeling played an important role in the evolution of ideas about emotion. Our intuitive guess, emerging out of a commonsensical understanding of ourselves and the world,

would be that the feeling comes first, with the bodily changes following in its wake. But an eminent nineteenth-century American psychologist, William James, seemed to think otherwise. When James published an article titled “What is an Emotion?” in 1884, he unwittingly fired the first shot in a long-drawn battle among several competing theories of emotion. He asks the question more pointedly: do we run from a predator because we are afraid, or does the act of running produce fear? What comes first—the feeling of fear, or the bodily response? James proposed that the feeling is a result of bodily response. To state his proposal in his own words:

My theory ... is that the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion. Commonsense says, we lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are insulted by a rival, are angry and strike. The hypothesis here to be defended says that this order of sequence is incorrect ... and that the more rational statement is that we feel sorry because we cry, angry because we strike, afraid because we tremble ... Without the bodily states following on the perception, the latter would be purely cognitive in form, pale, colorless, destitute of emotional warmth. We might then see the bear, and judge it best to run, receive the insult and deem it right to strike, but we should not actually feel afraid or angry.

There are two possible accounts of the cause and effect relationships of an emotional response:

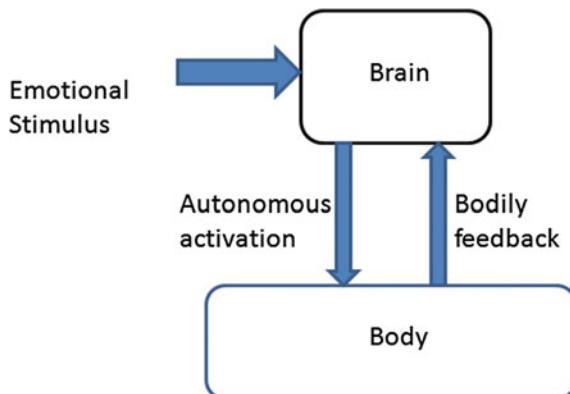
- (1) Stimulus (predator) → feeling (fear) → bodily response (running)
- (2) Stimulus (predator) → bodily response (running) → feeling (fear)

James chose option #2 over what seems acceptable to common knowledge, option #1. James’ theory is primarily emphasizing the importance of bodily or physiological response to emotional experience. Without the feedback of physiological response, emotional experience would be bland, placid, and incomplete. Each type of emotion would be accompanied by physiological responses that are distinct and unique to that emotion. A votary lost in her rapture of God may choke and shed tears of joy but is not likely to run, while an individual under an attack by a predator takes to his heels. It is this distinctness in bodily response that gives the emotion its distinct color.

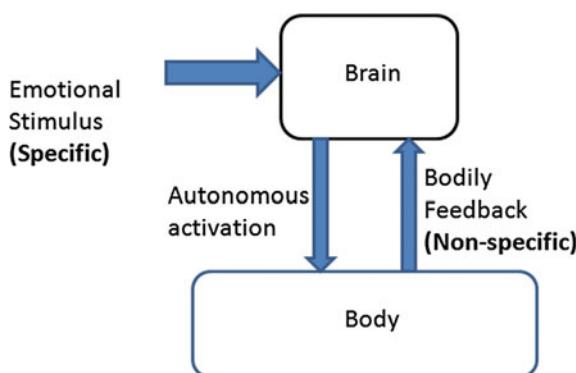
Carl Lange, a Danish physician, developed a theory of emotions that closely resonates with ideas of James. Like James, Lange also believed that emotions are caused by physiological responses. But unlike James, he emphasized the specific role of vasomotor responses. Therefore, the theory that emotional feeling is caused by the physiological response is referred to as James–Lange theory (Fig. 10.2).

James’s ideas received wide acceptance by psychology community for several decades. However, in 1920s, first notes of dissent began to be heard. One of the first to oppose James’ ideas was William Cannon, a physiologist at Harvard medical school. Cannon developed the idea of homeostasis originally proposed by Claude Bernard, a French physiologist. Claude Bernard proposed that the internal environment of the body is actively maintained at constant conditions by the action of nervous system. We now know that the autonomic branch of the nervous system is mainly responsible for maintenance of such internal constancy. Cannon elaborated this idea and studied the nervous mechanisms underlying homeostasis. He coined the term

**Fig. 10.2** James–Lange theory of emotion. Emotional stimulus received by the brain produces autonomous activation in the body, which when fed back to the brain causes emotional experience



**Fig. 10.3** Cannon–Bard theory pointed out that the bodily feedback is the same irrespective of emotional stimulus, thereby dealing a deadly blow to James–Lange theory



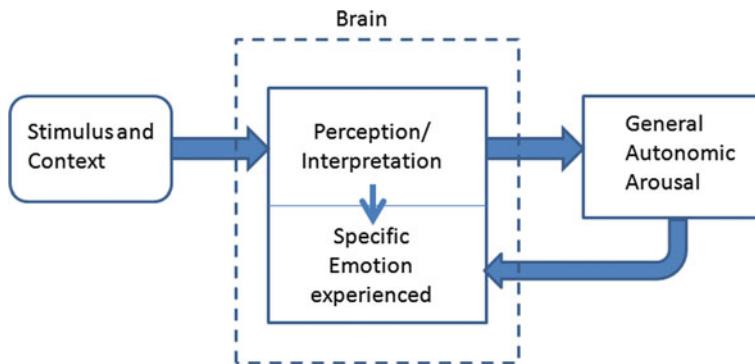
fight-or-flight, mentioned above, which refers to the response of an organism under attack. Cannon believed that the fight-or-flight response is mediated by the autonomic nervous system, particularly the sympathetic nervous system. A peculiarity of the sympathetic nervous system is that it responds in a general way, irrespective of the stimulus, a property known as mass discharge. No matter what the nature of the stimulus is, sympathetic activation produces a full-blown response involving accelerated heart rate, perspiration, piloerection, and so on. Cannon noted that independent of the type of the emotional experience, the bodily, sympathetic response is the same. This observation struck a serious blow to James–Lange theory which depended on the physiological response to provide specificity to emotional experience. Cannon felt that the factors that give specificity to emotional response must be sought after in the brain, or in the parts of the brain that receive the feedback from ongoing physiological responses in the body; they cannot be found in the bodily responses themselves. Cannon developed these ideas jointly with physiologist Phillip Bard. The resulting theory is called Cannon–Bard theory of emotions (Fig. 10.3).

A resolution of the standoff between James–Lange and Cannon–Bard theories of emotion came much later in the '60s. Part of the reason behind the delay was the

behaviorist thinking that had a strong influence on psychology throughout a good part of twentieth century. Behaviorists held that all that an organism does must be expressed in terms of behavior and nothing else. Subjective elements like thoughts, feelings, and emotions are non-existent in behaviorist view. Insertion of these subjective notions in psychology was felt to be contrary to the objective standards of science. Prominent behaviorists like B. F. Skinner, Edward Thorndike, and John Watson rejected all introspective methods and resorted solely to experimental methods in which behavior can be clearly defined and quantified. It was perhaps necessary to take such a rigid stand on matters pertaining to the mind, since it was a time when appropriate tools for probing the brain were not available. In the absence of the right tools, researchers resorted to fuzzy speculation about mental processes marring the development of science. In such a setting, the question of the causes of emotional feeling was not considered a serious scientific question. A framework that refused to accommodate feelings in the first place naturally found the origins of feeling irrelevant. But a new wave began in the second half of twentieth century. The behaviorist movement started giving way to the cognitive revolution. The cognitivists sought to explain all mental functions in terms of cognitions, in terms of a well-defined sequence of internal processes by which an organism responds to an external stimulus. The question of causes of feeling, particularly the factors that are responsible to the specificity in emotional experience, began to be given fresh attention.

Two social psychologists at Columbia University, Stanley Schachter and Jerome Singer, set out to investigate the standoff in the two key rival theories of emotion. They discovered that each of the rival theories was partly true. The physiological response was indeed important just as James had suggested, but it lacked specificity just as Cannon pointed out. A range of emotional experiences is associated with a common set of physiological responses—sweaty palms, pounding heart, and so on. The heightened state of arousal is indeed essential for the intensity of emotional experience. But it turns out that specificity comes from somewhere else. The immediate external conditions, the social context, the stimulus, determine the specific emotion that is actually felt (Fig. 10.4). The pounding heart and sweaty hands could signal feelings of joy if what you have in front of you is a person that you love, and feelings of morbid fear if it is a hissing snake. Schachter and Singer set out test their hypothesis with an experiment.

The subjects in this experiment were kept in the dark regarding the ultimate objective of the experiment, to make sure that knowledge of the objective does not bias and color the emotional responses of the subjects. Subjects were told that the experiment was about testing the effects of vitamins on vision. Specifically, they were told that a vitamin compound called Suproxin was being tested. But actually some of the subjects were given a  $\frac{1}{2}$  cc dose of epinephrine, a drug that activates the sympathetic nervous system; the remaining subjects were given saline water as placebo. Those who were given epinephrine were further segregated into three groups: the “informed,” the “ignorant,” and the “misinformed” group. The “informed” group was told that the drug can have side effects like increased heart rate, thereby allowing them an explanation of the experiences they were going to have. The “ignorant” group were told nothing. The “misinformed” group was told that they were going to



**Fig. 10.4** Schachter and Singer theory of emotion: specific emotional experience is produced by a combination of cognitive interpretation of the specific emotional/sensory stimulus and the general autonomic arousal

have side effects that had no relation to those they were really going to experience. With the conditions of arousal of the subjects thus controlled, the experimenters also arranged for appropriate forms of emotionally significant stimuli. Two trained actors were engaged to act before the subjects in “euphoric” or “angry” manner. At the end of the experiment, the subjects were evaluated regarding the nature and intensity of their experience. Participants who were given the drug that produced sympathetic activation felt a greater sense of arousal than those who were given a placebo. Furthermore, among those who experienced such higher arousal, the ones who were exposed to “euphoric” display reported feelings of euphoria, and those exposed to “angry” displays had feelings of anger. Thus though the arousal was the same, the specific feelings depended on the external conditions. The experiment thus provided strong support to Schachter and Singer theory.

Let us take stock of where we stand at the moment in our search for a satisfactory theory of emotions. James–Lange’s proposal that emotional experience depends on the feedback from the body was countered by Cannon–Bard theory which pointed out that the autonomic state is nonspecific and therefore cannot account for specificity in emotions. Schachter and Singer theory confirmed parts of both and achieved some sort of reconciliation between the two theories by demonstrating that bodily feedback intensifies emotional experience but does not provide specificity, which seems to arise out of a cognitive interpretation of external stimuli and social context. If bodily feedback only intensified emotional experience, the latter must have arisen even before the autonomic response in the body had developed. This aspect of delay in autonomic responses was pointed out by Cannon also who noted that autonomic responses are too slow to account for emotional feelings. Therefore, it remains to be clarified, how and where do the feelings occur?

The second half of the twentieth century is an era of cognitive science and cognitivist thinking about mind and brain function. There was an attempt to resolve every known psychological process into its component steps and the sequence in which

those steps occur. A similar approach was directed toward study of emotions too. This led to the birth of appraisal theory of emotions, which posits that emotional feeling depends on cognitive appraisal or evaluation of the external conditions.

A strong proponent of the appraisal theory was Magda Arnold (1903–2002), a brilliant psychologist and a remarkable individual who lived for a ripe age of 99. Her major work on emotions was *Emotion and Personality*, a two-volume work that was published in 1960. In this work, she sought to study the brain circuitry that subserves perception, motivation, and emotion.

The broad idea behind linking emotion to cognitive evaluation may be expressed with an illustration. Consider the experience of facing a job interview. Assume that at the time of the interview you were in a financial tight spot and needed the job desperately. Consider that the first few brilliant answers impressed the interviewers and your job prospects began to glow. You could not contain the excitement and could feel your heart pounding away. But a sudden query thrown by a chap, who was sitting quietly at one end of the table until then, put you on the spot. As you scrounge the depths of your memory for an answer that is hard to find, your evaluations of the current situation start to nosedive. You get that sinking feeling and your tongue runs dry. But suddenly, by a stroke of luck, an answer flashed in your mind, and interview board found your reply so convincing that they ask you when is the soonest you can join. Your evaluations soar once again. The emotional roller coaster ride that you have been on during those few tens of minutes is predominantly steered by your interpretations and evaluations (“Is it good for me, or is it dismal?”) of your immediate situation. Thus, the emotion that is experienced depends on the cognitive appraisal or evaluation (is it good for me or bad?) of the immediate context.

Richard Lazarus, a psychologist at the University of Berkeley, took the appraisal theory of emotions further. Emotion, he argued, is the result of an appraisal that people make, about their immediate surroundings, about the situation they are facing, and the meaning of that situation to their well-being and survival. Lazarus also developed a theory of coping in which the appraisal is linked to stress. According to his theory, stress has more to do with how the subject felt about his resources than the subject’s actual situation. Thus, the appraisal, or an emotional evaluation, can become more important than the reality of the situation. Such a view of emotions shows the phenomenon of denial in a new light. Patients who are engaged in denial about their health condition are found to fare better than those who had a realistic assessment. Their favorable appraisal, even though removed from their reality, is helping them cope with their condition. In a book titled, *Stress, Appraisal and Coping*, co-authored with Susan Folkman, Lazarus explored the relationship with stress and coping. Effective appraisal, which leads to successful coping, helps you cope with stress. When the appraisal is ineffective, and coping fails, stress builds up and manifests in a range of pathological effects ranging from physiological disturbance and impaired social functioning.

Richard Lazarus described appraisals in terms of something positive or negative happening (or happened, or going to happen) to the concerned individual. For example, fear represents an immediate and overwhelming physical danger. Sadness represents an irrevocable loss that had already happened, whereas happiness repre-

sents a progress made toward a desirable goal. Disgust represents a state in which one is too close to an object or an idea that is difficult to assimilate. Such a treatment of emotions and their appraisal lends itself, as we will see in the subsequent sections, to creation of a more neurobiologically grounded theory of emotions.

Several others have followed the tradition of appraisal theory that is based on the premise that cognitive appraisals are the very essence of emotions. This cognitive component of emotions seems to make emotions easily analyzable. The theory maintains that through introspection, the results of which can be elicited by interrogation, it is possible to analyze and understand the contents of emotion. A study by appraisal theorists Craig Smith and Phoebe Ellsworth asks subjects to assess a past emotional experience in terms of several emotional dimensions (like pleasantness, effort involved, etc.). For example, pride may be associated with a situation involving pleasantness and little effort, while anger is linked to unpleasantness and a lot of effort. Thus, it seemed to be possible to resolve emotions to their bare essentials simply by introspection and verbal report of the same.

Thus, a cognitive approach to the problem of understanding emotions seemed to have achieved a tremendous success, paradoxically, by reducing emotions to a cognitive phenomenon. A major complaint that is often leveled against early cognitive science and its allied fields like artificial intelligence is that they have successfully banished emotions from their purview. A robot or a computer system with emotions is often seen a creation of science fiction, as a wonder that cognitive science and computer science can only aspire to create, but hitherto failed to achieve. In making emotions a part of the cognitive reckoning, the appraisal theorists seemed to have bent backward and committed an opposite error. By reducing emotions to verbal analyses and self-reports, they seemed to have expelled the charm, the intrigue, the sweet or terrible unpredictability, from emotions. If emotions can be seen and read out so clearly, like a piece of text under a table lamp, they would not be emotions in the first place. A good theory of emotions must allow them their right to be mysterious, at least in part. This mysterious aspect of emotions began to be unraveled through the link between emotions and the unconscious.

## The Unconscious Depths of Emotions

Consider the following striking experiment that highlights the irrelevance of cognition to emotional preference. Subjects were shown some emotionally neutral, non-textual patterns like the Chinese ideograms in two rounds. The set of patterns shown in the first round may be labeled as “previously exposed” and those shown in the second round as “novel.” The patterns—old and new—are jumbled and presented to the subjects who were then asked to choose the patterns they find more preferable. The subjects chose some but could not rationalize their decision. It turned out that the subjects chose the ones they were “previously exposed” to. But in a given mixed set, the subjects could not tell the two sets apart. They only knew subconsciously that the first set was preferable over the second.

This is a simple instance where conscious cognitive appraisal was not involved in the formation of emotional responses. There was an underlying reason behind the subjects' choices which the experimenter knew, but was hidden from the view of their cognitive appraisal. Experiments like this one and many others were performed by Robert Zajonc, a social psychologist, to show that emotional appraisal need not be cognitive. The reasons of an emotional appraisal could be hidden from any cognitive reckoning. What shaped the preference was what was known as "subliminal exposure," an unconscious exposure to a stimulus.

A lot of experiments of the above kind were based on different ways of reaching the unconscious, bypassing the cognitive self. One way to do so in the visual domain is to present a visual stimulus so briefly that it fails to form a conscious registry. In one experiment, the subjects were shown some emotionally significant pictures (like a smiling or a frowning face), albeit too briefly (5 ms) to be registered consciously by the subjects. This pattern, known as the priming pattern, was followed by a masking stimulus which prevents the subjects from consciously recalling the original pattern. After a further delay, the target pattern, an emotionally neutral pattern, like the Chinese ideograms for example, was presented. The presentation of priming pattern → mask → target pattern was repeated over many patterns. The subjects were asked regarding the patterns they preferred. It turned out that they preferred those for which the primes had an emotionally positive significance (like a smile). But when probed regarding the reasons behind their choice, the subjects were unable to make their reasons explicit. Once again subliminally presented stimuli shaped emotional preferences without informing the conscious self.

This ability by which humans show evidence of perceiving a stimulus though they report no conscious awareness of that stimulus has been dubbed subliminal perception. Such perception has nothing to do with emotions per se. Earliest studies in this area date back to 1800s and early part of twentieth century. In these studies, for example, people were presented with visual stimuli from such a distance that it is nearly impossible to identify the stimuli. Or they were presented with auditory stimuli too weak to comprehend. They were then asked to identify the stimuli. In one such study, the subjects were presented with visual stimuli which could be single letters or single digits with equal probability, and were asked to guess the class (digit or letter) of the stimulus. The subjects reported that they were guessing but guessed at levels much higher than chance level.

Cases of subliminal perceptions have also been discovered in patients who underwent surgery under general anesthesia. As a matter of principle, general anesthesia is administered such that the patient is completely oblivious of what has happened during the surgery. This is often confirmed by checking with the patient once the patient is back to consciousness. The patient often denies remembering anything that happened during that time. But more delicate methods of probing have revealed that patient retained memories of stimuli presented under conditions of general anesthesia. For example, in one such study, the patients were played sounds of certain words (e.g., guilt, prove, etc.) repeatedly when under general anesthesia. Once they came back to consciousness, they were given word stubs like gui-, pro-, and so on

and asked to fill the blanks. The patients chose the words they “heard” under general anesthesia to fill the missing parts of the words.

There were claims that the power of subliminal perception was exploited by companies to influence consumers and induce them to buy their products. One such claim which was published in 1957 was made by James Vicary, a market researcher. Vicary described an experiment in which a large number of patrons were exposed to two advertising messages: “Eat popcorn” and “drink coco-cola” while they were watching a movie in a theater in New Jersey. According to Vicary’s report, the messages were flashed only for 3 ms, a duration too brief for conscious recognition. Over a 6-week period during which these messages were presented, the sales of popcorn rose by 57.7% and that of coke by 18.1%! Vicary’s study was described in a popular book titled “the hidden persuaders” by Vance Packard. The book described how companies manipulate the minds of consumers persuading them to buy their products, and how politicians use the same tactics to influence voting patterns. All this led to public outrage and resulted in creation of a law that prohibits use of subliminal perception in advertising.

Thus, the phenomenon of subliminal perception shows that conscious perception of a stimulus is not necessary to exhibit behavior that depends on the stimulus. Such stimuli can influence emotional preferences and decision-making, while completely evading conscious perception. Existence of subliminal perception turns out to be perhaps the strongest counterargument to appraisal theory of emotion. Conscious, cognitive appraisal cannot be the whole story with emotions. At the same time, AI-style approaches that hope to reduce emotional processes to a set of clean, well-defined procedures, and design machines that posses (cognitive!) emotions, are foredoomed. Subliminal perception strongly urges us to fundamentally rethink our strategy of understanding emotions.

There is an even more dramatic form of sensory phenomenon in which the subject denies any conscious experience of the stimulus but exhibits behavior that depends on reception and processing of the stimulus. A class of patients who suffer from a condition called blindsight report no visual awareness but are capable of displaying visually based behavior. One of the earliest patients of blindsight was DB, a patient at National Hospital at London. DB had his occipital lobe surgically removed as a treatment to remove a tumor that invaded it. Although DB reported that he could not see anything, he could perform a lot of visual tasks. For example, he could reach out to objects with remarkable accuracy. When shown a grating pattern, he could tell if the lines are oriented one way or the other. Since the early studies on DB, a large number of blindsight subjects have been studied confirming the phenomenon. Analogous conditions in touch (“numbsense”) and even hearing (“deaf hearing”) have also been found. This unconscious sensory capacity, which occurs when the corresponding sensory cortex is damaged, is believed to be possible because certain relevant deep brain structures, like thalamus, for example, are intact. Since most sensory information goes to thalamus before it proceeds to cortex, it is likely that thalamic representations of sensory inputs are responsible for this kind of unconscious sensory capacity.

Let us pause for a moment to take bearings on our journey through various influential theories of emotions. We began with James–Lange theory that emphasizes the importance of bodily response for emotional experience. But Cannon–Bard theory points out that the bodily response occurs rather late, and also argues that it lacks the specificity required for various emotions. Schachter and Singer theory reconciled the two theories by proving that though bodily response is important (it intensifies emotional experience), specificity does not come from it. Specificity comes from a conscious evaluation and interpretation of external conditions and social context. This paved way to appraisal theory and the thinking shifted in a major way from bodily response to cognitive evaluations. Subliminal perception and related results brought about correction of a different kind. Between the cognitive self and the bodily self, it posited an unconscious self that can influence and determine the contents of emotions. A greater synthesis, with James–Lange approach at one end of the spectrum, and that of the appraisal theorists at the other, seems to be the need of the hour.

In our quest to understand emotions, we seem to have got ourselves stuck in the body–unconscious–cognition axis. If a grand synthesis is our ultimate goal, we are not likely to succeed merely by collecting more data that support various theories located somewhere on this axis. We need to get out of this axis in the first place and search in a very different direction. Indeed, there is a direction, a line of attack on the problem of emotions that we have ignored in the story of emotions that we have narrated so far. The theory, or the spectrum of theories of emotions that we have so far visited, is predominantly psychological. Beyond the initial references to involvement of autonomic responses in emotion, our considerations have been, almost exclusively psychological, not really neural. The influence of behaviorism led us away from drives and motivation and other ghosts in the brain. The influence of cognitive science taught us the language of appraisal, and a strong reliance on introspection, from the point of view of the subject, and verbal report, from the point of view of the experimenter. The physical brain, the ganglia, the nuclei, the neurons, and the gossamer wiring that links them did not become a part of the reckoning. A theory of emotions that does not mention brain is not likely to get very far.

Part of the reason behind this gross omission is perhaps the subconscious (?) feeling that emotions are an exclusive prerogative of humans, conscious and intelligent, and cannot be unraveled by any means other than linguistic, cognitive investigation, and analysis. The legend and folklore that surrounds the claims of pet owners about the boundless love that their pets shower on them is not often sufficiently convincing to the scientifically inclined. But it is an unmistakable sign of vanity to deny the entire animal world their share in emotions, when we ourselves are a precarious outgrowth of aeonic animal evolution. Therefore, it seems to be eminently logical to reconsider the problem of emotions from two points of view: (1) from that of the brain and nervous system, and search for substrates of emotional activity in the brain, and (2) that of emotions as they occur in animals. Though a complete and comprehensive neural theory of emotions does not exist as yet, we will see that a familiarity with neural components of emotional activity offers some solid leads toward such grand synthesis.

## Animal Emotions and Facial Expressions

Earlier in this chapter, we have seen how Sylvan Tompkins, Paul Ekman, and others have classified emotions purely on the basis of facial expressions. We have mentioned this class of studies as some of the first in line that led to the development of psychological theories of emotion. But there is a study of emotions based on facial expressions, one that is much older than that of Sylvan Tompkins, and one that is perhaps the first of its kind. This study performed by Charles Darwin considered facial expressions in both humans and animals, and showed universal similarities across a range of species. An approach to emotions that discusses them in both humans and animals in a single breath necessarily paves way to a whole different direction of thinking, a direction that would perforce lead to a neurobiology of emotions.

But Darwin's original question was not to study emotions, human or animal, but to understand the forces that shape the evolution of species. Rejecting the religious idea that God had created different species afresh at various stages in the history of the Earth, Darwin set out on a worldwide voyage, famously called the "Voyage of the Beagle," to look for evidence. He brought back with him a large mass of biological evidence, in the form of specimens like teeth and feathers, bones and nails, and fossils of all types. His studies convinced him that biological organisms were not created arbitrarily afresh without a precedent, but have evolved gradually through the history of the Earth. Certain biological traits are inherited from generation to generation. Children do resemble their parents and often inherit certain traits. But certain new traits also emerge in new generations, traits that did not exist before. Children do not look identical to their parents. Sexual reproduction is not only a preserver of old traits, it is a source of new ones. Thus, Darwin observed that change is a natural process of evolution. But what is the logic behind this change? What is the direction of its progress? In answer to this question, Darwin thought of how animal breeders tend to pair certain breeds together in order to produce offspring with suitable traits. The criteria for interbreeding that determine the choices of a breeder include more milk, more meat, more speed and strength, greater resistance to disease, or simply longer life. Thus optimizing certain traits is the direction in which the controlled evolution shaped by a breeder proceeds. By extension of this idea to nature at large, Darwin visualized that species compete for the limited natural resources in a struggle to survive. Thus, species that are fit in a given natural habitat tend to survive, passing on their traits to their offspring. Mother Nature, like a grand Breeder, selects the species that are the fittest for survival in a given milieu. This "natural selection" is the cornerstone of Darwin's theory of evolution.

But Darwin felt that not only physical features but personality traits and behavioral features could also be inherited. Particularly, Darwin observed that there are universal ways of emotional expression, through countenance and bodily gestures, that cut across cultures and species. He captured his observations on this subject in a volume titled "*The Expression of the Emotions in Man and Animals*," in which he expounds three principles of expression:

The first principle, dubbed the principle of serviceable associated habits, states that certain bodily movements or actions are habitually associated with certain states of mind. Through repeated association between the states of mind and accompanying actions, expression becomes habitual, and sometimes even outlives its original purpose.

The second principle, known as the principle of antithesis, may be simply stated as follows. If a state of mind #1 is exactly opposite in nature to state of mind #2, the expression of the former will be exactly opposite to those of the latter.

The third principle which reads, “*The principle of actions due to the constitution of the Nervous System, independently from the first of the Will, and independently to a certain extent of Habit,*” needs some translation since he uses a rather archaic language, very different from that of contemporary neuroscience. Here, he talks about certain reflexive, innate motor traits that are hard-wired in the nervous system, and have nothing to do with the action of the will, or formation of habit.

Darwin gives a large number of examples to support these principles. A perplexed man tends to scratch his head as if the external act is going to provide relief to his internal crisis. Accordance with another's view is expressed by a nod, while opposition is expressed by shaking the head. A person describing something horrible shuts his/her eyes, or shake his/her head, as if trying “not to see or to drive away something disagreeable.” A person trying to recollect something tends to raise his/her eyebrows attempting actually to look for it. Interestingly, in this context, Darwin comments that “a Hindoo gentleman,” a person of Indian origin, also agrees that the trait of lifting eyebrows during mental recall is common even in the subcontinent.

Darwin also observes that certain motor traits are inborn even among animals. Young ponies show inherited gait patterns, like ambling and cantering. A certain moth species (humming-bird Sphinx moth), soon after its emergence from its cocoon, is seen to hover above a flower, inserting its proboscis into the orifices of the flower. Certain species of pigeons display peculiar patterns of flight which could not have been learnt. In addition to such general motor traits, Darwin also observes innate patterns of expression of emotion. In face of a mortal threat, it is common among a variety of animals, and also humans, to urinate and defecate. Piloerection is a common form of emotional expression found in many animal species including rats, bats, cats, lions, and so on. A remnant of this primordial expression in humans is goosebumps, a manner of emotional expression that could indicate an elated state of mind, or one of sheer horrors. A state of rage that is accompanied by snarling, hissing, spitting, baring of canines, and other gestures is found in both animals and humans. Quoting from Shakespeare, Darwin presents an illustration of how an agitated state of mind could be expressed in an elaborate, quaint bodily ritual:

Some strange commotion  
Is in his brain; he bites his lip and starts;  
Stops on a sudden, looks upon the ground,  
Then, lays his finger on his temple: straight,  
Springs out into fast gait; then, stops again,  
Strikes his breast hard; and anon, he casts



**Fig. 10.5** Similarity in human and simian facial expressions

His eye against the moon: in most strange postures

We have seen him set himself.—*Hen. VIII.*, act 3, sc. 2.

The observation that humans and animals share certain forms of emotional expression urges to look for common origins (Fig. 10.5). We need to figure out methods of emotion research that could be applied equally to humans and animals. The methods described hitherto were predominantly cognitive, with a strong dependence on verbal report and introspection. These psychological methods were certainly fruitful in case of human emotions, but will be irrelevant to unravel animal emotions. If traits of emotional expression were inherited by us from our animal past, the basis of that inheritance can only be found in our brain which has evolved from a mammalian brain, or a primate brain, more specifically, and look for correspondences between shared cerebral features and shared manner of emotional expression. Even if we are ambivalent in granting animals the luxury of subjective emotions, we have no option but to allow them emotions in their objective expression. A new line of investigation into the nature of emotions now opens up. This line of study will have to consider a neural basis for emotional expression. After romancing with feelings, appraisals, conscious and unconscious, and cognitive evaluations for too long, we realize it is now time to return to the bodily basis of emotions. And when we consider the body, the brain will follow in its wake automatically, autonomically.

## Emotions Right in the Middle

Some of the earliest efforts to find the “centers of emotion” in the brain came from a Russian physiologist named Vladimir Bekhterev. Conducting experiments on animals in which parts of the brain were ablated, Bekhterev looked for signs of emotional attenuation. As a result of such experiments, in 1857, Bekhterev concluded if the brain is lesioned or “truncated” above an important structure called thalamus,

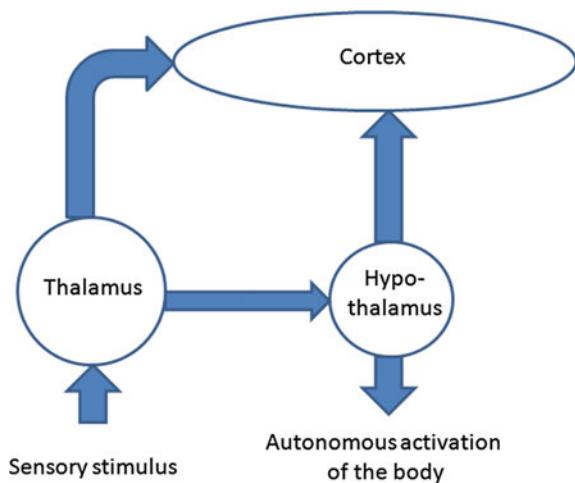
emotional responses remained nearly intact. Thalamus is an important hub through which most sensory data (except olfactory and gustatory information) passes, before it arrives at appropriate targets in the cortex. This strategic anatomical location of thalamus had earned this structure the title “gateway to the cortex.” Bekhterev concluded that thalamus is the center of emotion. But this line of work went unnoticed for nearly three decades. In 1892, Friedrich Goltz followed up Bekhterev’s work. Goltz was interested in studying the contributions of the cortex to motor output. He studied decorticated animals, a special neuroanatomical preparation in which the entire cortex is disconnected surgically, and found that they exhibited a lot of normal functions of nervous system like sleep–wake cycle, temperature rhythms, spinal reflexes, response to sensory stimuli, and so on. Most importantly, they showed normal responses to painful stimuli. Though it was not possible to attribute experience of pain to these animals, they obviously exhibited reflexes of pain. Although the use of decorticated animals could have become a fruitful line of work in emotion research, unfortunately this tradition was ignored for some time when Walter Cannon entered the scene.

While decorticated animals were continued to be omitted, emotion researchers often used a so-called “cat and dog” paradigm in experimental studies of emotion. In this setup, the cat is placed inside a cage and the dog is placed outside. The dog is angry and frustrated that it is prevented from reaching the cat. The cat is afraid and exhibited signs of fear response. Cannon found this setup inconvenient for careful physiological measurement. When the animals are in such an excited condition, it is difficult to measure circulatory and other autonomous responses like blood pressure, temperature changes, etc. Measurement, taken when the animals become calm again, defeat the original purpose. Cannon found that the decorticated preparation is full of potential for understanding substrates of emotions in the brain.

One of the first things that Cannon noticed about decorticated animals is that, under provocation, these animals showed a range of responses bundled as rage response consisting of retraction of ears, unsheathing of claws, arching of the back, growling, and so on. This evidence flew in face of James–Lange theory which maintained that cortex is necessary to process sensory information and produce motor output appropriate for emotional response. Cannon felt the need to distinguish this rage response from the one that occurs in an intact animal and coined a new term to describe the rage expressed by a decorticated animal—“sham rage.” Cannon was joined by his student Phillip Bard in his efforts to precisely establish the substrates of emotion. Their joint efforts resulted in Cannon–Bard theory of emotion (Fig. 10.6). Together, they ablated brain at various levels and systematically descended toward the diencephalon—a brain region that consists of two key structures, thalamus, and hypothalamus. They noticed that rage response was significantly attenuated when parts of hypothalamus were lesioned. Animals in which hypothalamus was destroyed did show some aspects of rage response—snarling, crouching, hissing, unsheathing of claws, etc.—but the responses did not occur in a coordinated fashion as they occurred in an animal with intact hypothalamus.

The crucial role of hypothalamus in coordinating rage and other emotionally significant responses have been unraveled through electrical stimulation experiment,

**Fig. 10.6** A schematic of Cannon–Bard theory of emotion processing. Hypothalamus is a key coordinating center for emotional expression; cortex is considered the site of emotional experience. In James–Lange theory, physiological response is the cause of emotional experiences. Contrarily, in Cannon–Bard theory, hypothalamus triggers both bodily response and emotional experience in the cortex



analogous to cortical stimulation experiments performed by Penfield. In experiments of this kind, electrodes were placed at various points in the brain of a caged experimental animal. The animal is free to press a lever that is placed next to the cage. When the lever is pressed, a mild current is delivered to the animal's brain through the implanted electrode. Stimulation of certain sites in the brain caused such a sense of reward or pleasure that the animal depressed the lever hundreds of times in an hour. Interestingly, the animal found this experience of self-stimulation preferable even over offerings of food. Sites of this type were found in parts of hypothalamus, the lateral and ventromedial nuclei.

Other sites were found in hypothalamus where stimulation produced a strong aversive response in the animal, as if the stimulus were punitive. Stimulation here elicited responses of pain, terror, and sickness. A rage response was characteristically elicited by stimulation of punishment centers, exhibiting hissing, spitting, crouching, snarling, and other classic features of rage. Such "punishment centers" were found particularly in the periventricular parts of hypothalamus, the parts that overlie ventricles, and also thalamus. It is noteworthy that when both reward and punishment centers were stimulated, activation of the punishment centers inhibited the reward centers, and there was a predominance of fear and pain response.

The fact that hypothalamus is an important control center for coordinating emotional responses, the fact that it receives inputs from thalamus, and the fact that thalamus is involved in unconscious sensory processing open a window of opportunity that could possibly reconcile the standoff between James–Lange theory and Cannon–Bard theory. It could also begin to conveniently accommodate both the cognitive appraisal approach to emotions and the unconscious aspect of that appraisal. When sensory input winds its way through the nerves, it first arrives at the thalamus and proceeds onward to the sensory cortex, where it produces the corresponding sensory awareness. A part of the information that arrives at the thalamus is also

conveyed to hypothalamus which evaluates the emotional or affective significance of the sensory information and triggers appropriate bodily responses. The hypothalamus in turn projects to the cortex. Thus, the cortex is now a proud recipient of two streams—the direct stream from thalamus which conveys sensory information, and the indirect stream via the hypothalamus which supplies the emotional appraisal. The combination of these two forms of information in the cortex is probably the precondition for the feeling that goes with an emotional response.

This simplified depiction of emotional processing as it is played out in the circuit consisting of thalamus, hypothalamus, and the cortex has several merits. It agrees with James–Lange theory in part, in that it does not say that conscious experience is responsible for bodily responses. For in the above description, it is the unconscious processing occurring at the level of the diencephalon (thalamus + hypothalamus) that initiated the bodily response. It agrees in part with Cannon–Bard theory, since it does not invoke the feedback from the body, to account for emotional experience. In fact, the above picture does not even include such feedback. Next, it accommodates cognitive appraisal, since such appraisal can be said to be occurring in hypothalamus which has access to the sensory information available in thalamus. But it also stresses the unconscious aspect of such appraisal since the processing occurs at the level of thalamus and hypothalamus.

But hypothalamus is not the whole story of emotions. It is a good place to begin with, and is indeed a key player. But there is a whole world of brain structures, cortical and subcortical, that are involved in emotional processing. Let us begin with a brain area involved in emotions, an area that is right in the middle of the brain, and was a source a lot of progress and controversy in emotion research.

## The Middle Kingdom of Emotions

Imagine the shape that is produced when the fists formed by your two hands are brought together so that the middle phalanges of the four fingers come into contact. This shape that looks like a 3D object with two halves, or “hemispheres,” divisible right in the middle by a vertical plane that separates the two fists, has a convenient resemblance to the hemispheres of the brain. The visible parts of the surface of this double fist, the back of the hand and the lower phalanges of the fingers, are comparable to the lateral cortex of the brain, the part of the cortex that is visible from outside. The parts of the double fist that are in contact—the middle phalanges of the four fingers—are comparable to a cortical region that is located right in the middle of the brain, hidden from the external view. This part of the cortex, the medial cortex, has been named by the famous French neurologist Paul Broca as le grand lobe limbique, or the great limbic lobe in plain English. Broca thought of the limbic lobe as the fifth lobe, after frontal, parietal, temporal, and occipital lobes. Another reason he distinguished the limbic lobe from the other lobes is that this part of the brain is hardly convoluted, like the other four lobes. Since its appearance resembled the cortices of lower animals, he felt that limbic lobe corresponded to “bestial” nature

in us. Another reason behind attribution of “bestiality” to this cortical area is its link to sense of smell, which earned this brain area the title of rhinencephalon (meaning “smell brain”). Smell is an important driving force in an animal’s life and behavior. Smell is a strong guiding power in foraging for food, flight from a predator, and sexual arousal. Anatomist CJ Herrick, who did seminal work on the evolution of brain, felt, as Broca did, that the lateral cortex must be distinguished from the medial cortex, on evolutionary terms. He proposed that, whereas the older medial cortex is involved in emotion processing, the newer lateral cortex, called the neocortex (“neo” means new), is responsible for higher cognitive functions in humans.

There are other reasons for considering the association between the medial cortex, particularly a part of the medial cortex known as the anterior cingulate cortex, and emotions. An early case—and a royal one at that—of this link dates back to the seventeenth century. It involved a knight called Caspar Bonecurtius, who suffered from severe apathy. He sat in a place whole day long, unresponsive to his surroundings. Toward the end of his life, he spoke very little, and whatever little he spoke was not very meaningful. After his death, postmortem revealed a tumor in his anterior cingulated gyrus. Even in the middle of the last century, it was known that electrical stimulation or ablation of anterior cingulate cortex is accompanied by transient changes in emotional behavior. These findings led to a drastic practice of surgical removal of anterior cingulate in order to cure “severely disturbed mental hospital” patients. Damage to anterior cingulate also resulted in depression, apathy, delirium, and other affective disorders.

Another unlikely structure that was found to play a role in emotion processing was hippocampus. We have seen, in Chap. 5, that hippocampus is a site of memory consolidation. It is a place where memory resides temporarily, before it is shipped to long-term stores in the cortex. This mnemonic function does not give any clue to its role in emotions. But the link between hippocampus and emotions was first recognized from studies of cases of rabies. Rabies is a viral disease that affects the brain by causing inflammation. The symptoms may begin as headaches and fever, but expand to anxiety, insomnia, agitation, paranoia, terror, and consummating in delirium. The virus particularly attacks hippocampus.

Another disease—epilepsy—also links hippocampus to emotion processing. Epilepsy is a brain disorder characterized by seizures caused by uncontrolled spread of synchronized neural activity across the brain. These seizures are often preceded by an aura, a prior feeling, a sort of a warning sign that predicts a seizure. Interestingly, the auras are often accompanied by inexplicable fear, a sense of *déjà vu* (it happened before), and even a bad taste in the mouth.

Thus, a certain coarse knowledge of the above mentioned cerebral components of emotion processing was known even in the early decades of the twentieth century. Seizing upon these ideas, James Papez, an anatomist at Cornell University, made a bold attempt to expand the simple scheme of Cannon–Bard theory into a more elaborate circuit of emotions—the eponymous Papez circuit. The Cannon–Bard scheme primarily has two pathways: one proceeding directly from the thalamus to the cortex, and the other, a detour, that bifurcates from the thalamus and proceeds to the cortex but via an important hub of emotion processing—the hypothalamus. The essence of

these two branches is preserved in Papez circuit. Papez thought of these two branches carrying two fundamentally different streams of experience. The branch from thalamus to the cortex is thought to carry the stream of thought, while the detour from the thalamus to hypothalamus carried the stream of feeling. A broad distinctive feature of Papez circuit compared to the Cannon–Bard scheme is the presence of feedback from the cortex to hypothalamus; Cannon–Bard scheme only had a forward influence from hypothalamus to the cortex. These general differences and interpretations apart, what really put the Papez circuit on a pedestal is that it is primarily a neural circuit. Drawing from the available knowledge of the neural substrates of emotion at that time, he linked some specific neural structures in a circuit and proposed it as an engine of emotions. The new structures he added to Cannon–Bard scheme are hippocampus and cingulate cortex, for reasons mentioned above. Let us quote Papez himself on how he thought this circuit functions:

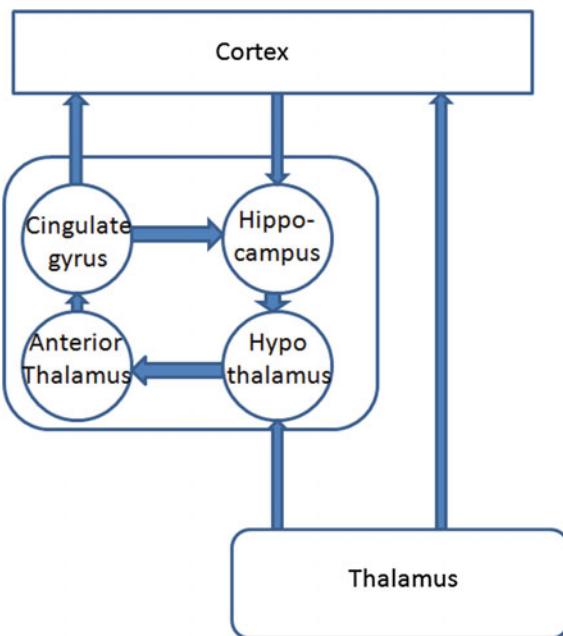
The central emotional process of cortical origin may then be conceived as being built up in the hippocampal formation and as being transferred to the mamillary body and thence through the anterior thalamic nuclei to the cortex of the gyrus cinguli. The cortex of the cingular gyrus may be looked upon as the receptive region of the experiencing of emotions as the result of impulses coming from the hypothalamic region, the same way as the area striata is considered as the receiver of photic excitations coming from the retina. Radiations of the emotive processes from the gyrus cinguli to other regions of the cerebral cortex would add emotional coloring to the psychic processes occurring elsewhere. This circuit may explain how emotional processes may occur in two ways: as a result of psychic activity and as a result of hypothalamic activity.

In Papez' view, the cingulate cortex is like a window between the cortex in general and the emotional circuit under the hood. Activity in the cingulate cortex, triggered by inputs from the hypothalamus (via anterior thalamic nuclei), spreads to other cortical areas. These inputs from cingulate cortex to sensory-motor cortex give the ongoing experience there an emotional color. Hypothalamic influences downward to the endocrine system and the autonomous system, produce bodily responses as in Cannon–Bard theory. A new element of Papez circuit, compared to the Cannon–Bard scheme is the feedback from the cingulate cortex to hypothalamus, via hippocampus. Thus, by supplying the neurobiological substance to Cannon–Bard approach, the Papez circuit had a profound influence on our understanding of emotion processing. But Papez circuit had an important emotion module missing. Ironically, some of the earliest work on this new module was published in 1937, the year in which Papez first published ideas about his circuit of emotions (Fig. 10.7).

## Almond Fears

In the '30s, Heinrich Kluwer, a professor at the University of Chicago, was studying visual cognition in monkeys. He was particularly interested in the effect of a hallucinogen called mescaline. As a part of his research, he tried the drug on himself and described some of his findings in a little book called *Mescal, the Divine Plant*.

**Fig. 10.7** Papez' circuit of emotions



Kluwer noticed that hallucinations that form part of the aura experienced by patients undergoing temporal lobe seizures resembled the hallucinations induced by mescaline. So he wanted to see if mescaline acted on the temporal lobe to induce those seizures. To check this idea, he wanted to remove temporal lobes in experimental animals and see if mescaline's ability to induce hallucinations will be blocked by the surgery. With the help of neurosurgeon Paul Bucy, he got this surgery—called temporal lobotomy—performed on monkeys. The results of the experiment turned out to be negative: the monkeys with temporal lobes removed continued to respond to mescaline as normal monkeys did. But what was interesting about the experiment was that the lobotomized monkeys showed some strange behavioral changes, which were summarized by Kluwer as follows:

1. Psychic blindness or visual agnosia: the ability to recognize and detect the meaning of objects on visual criteria alone seems to be lost although the animal exhibits no or at least no gross defects in the ability to discriminate visually.
2. ... strong oral tendencies in the sense that the monkey insists on examining all objects by mouth.
3. hypermetamorphosis: There is a tendency to attend and react to every visual stimulus.
4. profound changes in emotional behavior, and there may even be a complete loss of emotional responses in the sense that... anger and fear are not exhibited. All expressions of emotions... may be completely lost.

5. ...striking increase in amount and diversity of sexual behavior. It exhibits forms of autosexual, heterosexual or homosexual behavior rarely or never seen in normal monkeys.
6. a remarkable change in dietary habits... tendency to consume large quantities of ...[meat].

The above-listed complex set of symptoms was bagged together and described as Kluwer–Bucy syndrome. Both the cause (removal of temporal lobe) and the effect (the many-sided syndrome) are complex. The experiment does not provide a specific insight into the possible role of temporal lobe in emotional behavior, though it is clear that damage to temporal lobe indeed causes emotional disturbances. Therefore, efforts were on to localize the lesion further, and see if specific lesions could produce specific symptoms. As a part of this effort, Josephine Semmes from Yale and Kao Liang Chow from Harvard showed that lesions of temporal cortex produced “psychic blindness,” which refers to serious deficits in visual discrimination. Subsequently, Karl Pribram and Muriel Bagshaw showed that damage to deeper structures of temporal lobe, primarily the amygdala, produced the emotional changes (tameness, changes in sexuality, and eating) associated with Kluwer–Bucy syndrome. These findings put the spotlight on amygdala as an additional structure involved in emotion processing. Existence of amygdala was known even in the nineteenth century. The structure earns its name from the Greek word that denotes almond, referring to the almond-like shape of amygdala. But not much was known about amygdala’s function until the studies of Kluwer and Bucy and the research that followed. It remains to be seen how the new structure is related to the rest of the emotion circuit hitherto understood—the Papez circuit.

The story of our first efforts at gaining a functional understanding of amygdala begins with behavioral experiments on fear conditioning. In Chap. 1 of this book, we briefly encountered Russian physiologist Ivan Pavlov’s pioneering experiments on conditioning. The essence of this experiment was to learn a novel stimulus–response pair with the help of a pre-existing stimulus–response pair. When a plate of meat is offered to a hungry dog, the animal exhibits the natural response of salivation. In the terminology of conditioning literature, the plate of meat is an unconditioned stimulus (US). Before conditioning, the animal does not salivate in response to a neutral stimulus like the sound of a bell. But when the bell is rung a little before the meat is presented, the animal learns to respond by salivation even in response to the bell, which is now called a Conditioning Stimulus (CS). Salivation in response to CS is known as Conditioned Response (CR). Conditioning experiments are interesting since they represent simplest, and precisely quantifiable forms of learning, which are hoped to serve as prototypes to more sophisticated forms of learning that humans are capable of.

A variation of Pavlovian conditioning is fear conditioning, in which animals exhibit the so-called *fear response*, instead of salivation, in response to a painful or an aversive stimulus. A striking aspect of fear response is total immobility, as opposed to a more intuitive escape response in face of danger. This immobility or freezing may be thought of as a preparation for a subsequent escape, or a tempo-

rary strategy to prevent the predator from attacking, since predators often respond to movement. Fear response is associated with appropriate autonomous changes like increased heart rate and respiration, dilated pupils, etc.

In a typical fear conditioning experiment, a rat placed in a cage is first exposed to a sound followed by a mild electric shock. In the initial stages, the animal responds to the shock by freezing but ignores the sound. But when the sound and shock are repeatedly presented in that order, with a fixed delay, the animal begins to freeze in response to the sound. Joseph LeDoux and colleagues set out to unravel the circuitry that subserves fear conditioning. Their investigation naturally led them to amygdala which turns out to be an important hub in coordination of fear conditioning.

LeDoux and colleagues began their search for the branch, or the junction point at which the stream of sounds climbing up from the ear to the cortex meets the autonomous outflow that coordinates the fear response. The auditory stream begins its journey in the inner ear, where the vibrations produced by the sounds are converted into electric signals. These signals wind their way up toward the cortex passing several way stations, at various levels of the nervous system. First among these stations is the cochlear nucleus located in the brain stem, followed by inferior colliculus, another way station located slightly higher up in the midbrain. As the auditory information climbs further, it arrives at the thalamus, or more specifically the auditory thalamus, which refers to the thalamic nucleus responsible for receiving and relaying auditory information to the auditory cortex. Now what is the takeoff point on this auditory pathway at which a part of the auditory stream bifurcates and arrives at parts of the emotional circuitry that coordinates fear response?

Lesion experiments showed that damage to auditory cortex had no effect on fear conditioning, while damage to auditory thalamus or any way station below thalamus prevented fear conditioning. Thus, the auditory information must be branching out from the auditory thalamus to a target, other than the auditory cortex, through which it is coordinating fear responses. LeDoux and colleagues applied a classic neuroanatomical technique known as tract tracing to find out this new target. A question that neuroanatomists often find asking themselves is: does region A in the brain project to another region B? Or, conversely, does the region B receive inputs from region A? In an attempt to answer this question, a special visualizing substance called a tracer is injected into region A. The tracer then winds its way through the fibers connecting A to B. The wiring that connects B to A can then be visualized by standard staining methods. Similar methods applied to the auditory thalamus showed that this region projects to four different targets, one of which was amygdala. Which of these targets is responsible for fear conditioning? To answer this question, the research group systematically lesioned the four targets and checked for fear conditioning. Three of the targets had no effect on fear conditioning, but lesioning the fourth, amygdala, completely blocked fear responses.

We now know that a lesion of amygdala blocks fear conditioning. We also know how sensory information found its way to amygdala to trigger fear responses. But what exactly does amygdala do? How does it coordinate fear responses? A lot was known about the autonomic actions of amygdala long before the new direct connection between auditory thalamus and amygdala was discovered. Pioneering work

by Bruce Kapp and colleagues in 1979 unraveled the autonomic effects of activation of a central core of amygdala, known as the central nucleus. This central nucleus of amygdala has, as it was later worked out by several researchers, connections to hypothalamus and other brainstem areas by which it can produce autonomic responses like freezing, blood pressure, heart rate, etc. By selective lesioning of parts of the central nucleus, it was possible to block specific aspects of the fear response, for example, to eliminate increased heart rate, while retaining the freezing response.

We now have a concrete realization of the Cannon–Bard scheme applied to fear conditioning. Sensory input bifurcates at the level of thalamus into two pathways one proceeding to the sensory cortex, creating the sensory experience of the stimulus that is the original cause of the fear response. It is the sensory stream of Papez's depiction. The other branch from the thalamus proceeds to the amygdala where through specific outgoing pathways produces a whole array of autonomic changes that constitute fear response. This latter branch may be described as a part of what Papez visualized as the feeling stream. But unlike in Cannon–Bard scheme, it is not a direct projection from the thalamus to hypothalamus, but a direct thalamic projection to amygdala, that triggers the fear response. Thus, amygdala turns out to be the kingpin in the world of fear conditioning.

But a question that may be asked at this juncture is: what is the advantage of having two separate pathways, one for sensory experience and another for emotional response? In the words of Joseph LeDoux, why does the brain need the “high road” of the sensory pathway and the “low road” connecting thalamus and amygdala? First of all, is the auditory cortex even necessary for fear conditioning, since a copy of the auditory information is reaching amygdala through the “low road”? The answer is in the negative, since it was shown that tone–shock pairing could be achieved even without auditory cortex? Then what is the purpose of auditory cortex for fear conditioning?

In order to answer this question, Neil Schneidermann, Phil McCabe, and associates performed an experiment in which they tried to pair an auditory input that is more complex than a pure tone, with a shock. They presented two tones,  $T_1$  and  $T_2$ , with nearby frequencies, say,  $f_1$  and  $f_2$ . Only  $T_1$  was paired with the shock, but not  $T_2$ . The animal has to discriminate the two tones and exhibit fear response only to the appropriate tone. The animal was able to learn this more complex form of fear conditioning only when the auditory cortex was intact. When the cortex was lesioned, the animal exhibited fear conditioning to both the tones. This is because the information that travels down the “low road” does not have the detail that is characteristic of the information of cortical input. The two tones would sound nearly the same in the thalamus → amygdala pathway. The two sounds are discriminated at the level of the auditory cortex.

But our question is still unanswered. Why are there two pathways? If the auditory cortex is more informative, why not get rid of the “low road” completely? For one, the low road consisting of the projection from thalamus to amygdala is much older, in evolutionary terms, than the neocortex. So it is a baggage inherited from lower rungs of evolution, going all the way to reptiles. The advantage of this lower path is speed. It takes only a few tens of milliseconds at the worst for auditory information

to reach amygdala by the lower pathway. But it takes a few hundred milliseconds for the sound to be consciously registered in the auditory cortex. By the time the subject consciously perceives and identifies the auditory stimulus, the autonomous response triggered by amygdala would be well underway. By its very nature, a fear response is associated with an emergency situation, and rapidity of response is crucial. Therefore, evolutionary wisdom seems to have decided that it is better to act sooner, even if the action is based on coarse, approximate information, rather than opt for a leisurely response driven by conscious experience.

## Memorizing Fear

We have seen in Chap. 5 that damage to temporal lobe, particularly hippocampus, can cause serious memory impairments. The famous amnesic patient HM, who had undergone temporal lobectomy, had not only lost a good portion of his past memories (retrograde amnesia), he also had a difficulty in creating new ones (anterograde amnesia). The hippocampus is endowed with special neural infrastructure and the neurochemical mechanisms that enable this structure's memory-related operations. The kind of memory that hippocampus supports is known as declarative memory, a form of memory that can be consciously stored and retrieved. This form of memory must be distinguished from another memory system, the procedural memory, which refers to memory of motor skills. Procedural memories cannot be expressed or "declared" but are memories of nonconscious, implicit skills. This latter form of memory is subserved by a very different circuit known as basal ganglia. Thus, we have two parallel memory systems supported by apparently unrelated brain networks. But a closer study of damage to temporal lobe structures and the associated impairment of memory operations had unearthed a third form of memory, one that is also unconscious and had something to do with memory of emotions.

Edouard Claparede was a Swiss physician and child psychologist who lived in the earlier half of twentieth century. Claparede had an interesting amnesic patient who had retained some of the older memories but had lost all ability to create new ones. So all her experiences were short-lived and were erased within minutes. Claparede greeted her every day, afresh, while the patient never had a recollection of having met the doctor. As this ritual went on for some time, one day, when Claparede went to meet her, he extended his hand, as always, to greet her but with a pin concealed in his hand. The next day when he went to meet his patient, she again did not have a recollection of having met him, but simply refused to shake his hand. Claparede inferred that though his patient did not have the ability to remember conscious memories, she retained a memory of painful experiences, and of the fear response that the pinprick had elicited.

Not much was known in the days of Claparede about the neural substrates of this new emotional memory, a memory of pain and its consequential fear. But as knowledge of amygdala and its role in fear conditioning began to be accumulated toward the end of the last century, the observations pertaining to Claparede's patient seemed

to make more sense. This patient, like HM, must have had a damaged hippocampus, which explains her amnesia of declarative kind. But perhaps her amygdala was intact, which allowed her to store memory of a painful experience. It is interesting that the patient had no conscious understanding of why she hesitated to shake hands with her doctor. It indicates that the fear memory, which was supported by amygdala, was an unconscious memory. Thus, we have here a third memory system in addition to the declarative and procedural types, the one subserved by amygdala. This last type of memory is an emotional memory.

Now if we look back at our fear conditioning experiment from the previous section, we may come to regard conditioning also as memory. The rat had retained the memory that the CS (bell) is associated with a painful consequence, not very different from the manner in which Claparede's patient remembered (unconsciously) that the seemingly harmless handshake actually had a painful consequence. But the memories supported by hippocampus and amygdala seem to be of a very different kind—one retains memories of words, events, and other explicit items, while the other retains an unconscious memory of painful experiences. Considering the close contiguity of amygdala and hippocampus in temporal lobe, is it possible that the two memory systems are aspects of a larger memory system?

We have ignored an interesting feature in our earlier accounts of fear conditioning in rats. In these experiments, when a neutral stimulus like a tone (CS) is paired repeatedly with a shock (US), the rat learns to show fear response to the CS. But another element can also enter the picture and can trigger fear response in the animal. In addition to the CS, the cage, the surroundings in which the conditioning experiment was conducted, can by itself act as a trigger that can precipitate fear response. After sufficient training, if the rat is brought back to the same cage where it was earlier trained, it immediately shows signs of fear response (freezing, increased heart rate, etc.) without the necessity of presenting the CS. This form of conditioning is called contextual conditioning, since the surroundings or the context serves as a kind of CS in this case.

Therefore, there are two factors that contribute to fear response—the CS and the context. One may wonder why the animal's nervous system chose to split the environmental events into the CS and the context, since both may be thought of as parts of a unitary environment in which the animal is situated. The animal is trying to figure out the events in its immediate vicinity that can predict the arrival of a painful occurrence. In this process, it is trying to isolate cause and effect relationships from its experience of the environment, and thereby construct a useful model of the world. A hallmark of a good model is economy of representation. If there is a specific neutral event that consistently predicts the subsequent occurrence of a painful event, the animal is wiser to specifically pair the neutral event with the painful event, while deemphasizing other surrounding stimuli. But when there is no such specific neutral event, then the animal is faced with a harder task of building a cause and effect model of whatever it has at hand—to treat the entire context as being predictive of the painful event. Therefore, it was observed that contextual fear conditioning is more prominent when there is no CS at all. For the same reason, to really test whether the animal is sufficiently conditioned to respond to the CS, the animal has to be moved

to novel surroundings, to a different looking cage perhaps, and the experiment must be repeated. Damage to amygdala was found to block both types of conditioning. The animal responded to neither the tone nor the cage. But damage to hippocampus was found to selectively block contextual fear conditioning.

The role of hippocampus in contextual fear conditioning was verified even in human experiments. In one experiment, human subjects were immersed in a virtual reality environment which provided the context. The subjects were actually exposed to two such contexts: Context+ and Context-. Context+ was paired with a shock which served as US, as in the case of animal experiments. Fear response was measured using changes in skin conductance, a measure known as Galvanic Skin Response (GSR) linked to sympathetic activation. Contextual fear conditioning was observed in case of Context+ which was paired with shock. The subjects' brains were scanned using functional Magnetic Resonance Imaging (fMRI) technique while they performed the experiment. fMRI measures neural activity indirectly by measuring blood flow changes associated with neural activity in the brain. fMRI scans indicated significantly higher activation of hippocampus and amygdala in case of Context+ relative to Context- condition.

## Brain Mechanisms of Pleasure

Psychologists may visualize complex, multi-hued palettes of emotions; art folks may quibble about the perfect list of fundamental emotions; philosophers may hypothesize existence of exotic and unearthly emotions beyond the scope of common human experience. But if we descend to the level of the humble neuron, with its spikes and ion channels, there are only two very mundane emotions—pain and pleasure, the positive and negative that form the bedrock of all experience. Stimuli that elicit pain, the aversive stimuli, which make us run away from them, induce in us fear and panic. Stimuli that create pleasure in us, the appetitive stimuli, which make us want more of them, induce in us a sense of reward. Whatever emotional hues that may be must be constructed out of these binary colors, and are ultimately rooted in the gray axis that extends between reward (white) and punishment (black).

We have encountered the neural systems of fear response in the last section. Let us now visit the brain's engines of pleasure and reward. In 1954, two researchers, James Olds and Peter Milner, at Canada's McGill University, performed brain stimulation experiments in rats. Experiments in which brains were electrically stimulated in order to understand the responses elicited by the stimulation were known much before the studies of Olds and Milner. But in the experiments of Olds and Milner, the animals were given an option to stimulate themselves. When the animals pressed a lever, tiny currents flowed through brain regions where the electrodes were implanted. The question that the researchers asked is: will the animals prefer to press the lever, or avoid it? The studies were conducted with the electrodes placed at various locations in the brain. It was found that when the electrodes were placed in two specific brain regions—the septum and nucleus accumbens—animals pressed the lever at a

whopping rate of about 2000 times an hour! Some animals chose this stimulation over food, at the risk of severe starvation. Unwittingly Olds and Milner have hit upon a pleasure center of the brain. Studies scattered over several decades after the original findings of Olds and Milner have unraveled several other centers of pleasure in the brain. Examples of such studies include the stimulation experiments, described earlier in this chapter, which found pleasure centers in the hypothalamus. Studies that searched for brain areas that respond to pleasure have converged on certain key “hotspots” which include deep brain areas like nucleus accumbens, ventral pallidum, and brain stem, and cortical areas like orbitofrontal cortex, cingulate cortex, medial prefrontal cortex, and insular cortex.

In addition to the abovementioned cortical and subcortical pleasure centers, there is a small subcortical pool of neurons in the mesencephalon, known as the Ventral Tegmental Area (VTA) which plays a pivotal role in brain’s pleasure processing. VTA has neurons that release a chemical called dopamine, a molecule that is so important for pleasure that it has been dubbed the “pleasure chemical.” The relevance of dopamine for pleasure processing was first discovered indirectly when effects of blockage of dopamine transmission were studied. The pleasurable effect of stimulation of pleasure centers was found to be severely attenuated when dopamine antagonists, chemicals that block transmission, were administered. Dopamine antagonists were also found to attenuate the pleasurable experience that goes with addictive drugs like cocaine. Subsequently, it was found that both electrical stimulation of pleasure centers and addictive drugs stimulate neurons of mesencephalic dopamine centers. Not unexpectedly a more common desirable object like food also activated dopamine neurons. Application of dopamine antagonists attenuated this response to food stimuli also. These findings amply justify the title of “pleasure chemical” given to dopamine.

The key role of dopamine centers in pleasure or reward processing became clearer when anatomical investigations found that mesencephalic dopamine neurons project most of the other cortical and subcortical players of pleasure processing that we have listed above. Thus, it appears that the dopamine centers form the hub of the wheel of brain’s pleasure system. In addition to extreme or laboratory inducers of pleasure like electrical stimulation, or addictive drugs, and the more common, primitive rewards like food, brain’s pleasure system was found to respond to subtler forms of pleasure also.

The sight of a beautiful face is a source of pleasure, a fact that is used extensively in film, media, entertainment, and advertisement industry. Data from labor markets suggest that attractive individuals are more likely to get hired, promoted, and even paid more. In the ancient world, the influence of beautiful faces seems to have gone far beyond salary amplification, as was described potently by the Greek poet Homer when he wrote of “a face that launched a thousand ships.” Homer was singing of the disastrous graciousness of Helen of Troy, whose beauty precipitated the Trojan war. Functional MRI scans of people watching pictures of beautiful faces unraveled the secret of this ancient power: the pictures activated the reward system of the brain, particularly nucleus accumbens and VTA.

The sighting of a beautiful face can, if certain favorable conditions prevail, lead to romantic love, courtship and, if more favorable conditions prevail, to marriage. Based on a survey of 166 modern societies, it was found that romantic love is present in 147 cultures. The negative result obtained in case of the remaining 19 cultures, it was found in retrospect, was because the survey did not ask appropriate questions in those cases. Thus, romantic love seems to be a universal phenomenon with probable neurobiological bases. To test how brains respond to romantic love, Arthur Aron, Helen Fisher, and Lucy Brown took functional MRI scans of lovers. The subjects were shown pictures of their partners and some other individuals with whom the subjects did not have a romantic relationship. One of the key brain areas that were activated when the pictures of romantic partners were shown was once again VTA. In addition, other centers in the brain's reward system—insula, putamen, and globus pallidus—were also activated. These findings strongly suggest that love is such a powerful motivational force probably because it activates the reward system of the brain.

Notwithstanding the popular claims of the beneficial effects of humor on health, and the unsubstantiated celebration of humor's medicinal properties by popular adages ("laughter is the best medicine"), it would be universally accepted that humor is pleasurable. How then does brain respond to humor? In a functional MRI study that aims to answer this question, the subjects were shown 49 cartoons which were rated previously by a separate group as funny and non-funny. Brain areas that were preferentially activated when the funny cartoons were shown include the dopamine cell network of the mesencephalon and nucleus accumbens. Once again something pleasurable is found to activate brain's reward system.

Money is one of the most potent pleasure inducers, a power that ancient Greeks deified as Mammon, a prince of Hell. Wolfram Schultz and colleagues set out to study the effect of money on the brain using functional imaging. The subjects were shown certain complex images some of which were "correct." The subjects were asked to respond to the correct ones by clicking a mouse button. The subjects found out what the "correct" images were by the response from the experimenter. When the subjects responded to "correct" images, the experimenter simply said "OK" or actually gave a monetary reward. The study found that brain's reward centers (orbitofrontal cortex and midbrain centers) were preferentially activated when the subjects received monetary reward relative to the case when they received a neutral "OK".

Thus, a large number of studies have unraveled how the brain responds to the many forms of pleasure or rewarding stimuli. But what does the brain *do* with these responses? How does it act upon them? Pure happiness, unhinged from all earthly cares, may be the holy grail of the poet and the philosopher, but the fact that brain's pleasure responses are found not just in poets and the philosophers but in the brains of the rest of us, and also in others perched on the lower rungs of the evolutionary ladder like rats and monkeys, shows that the brain might have some serious purpose for its responses to pleasure. And why should it not? Pleasure or a persistent form of the same, happiness, is a strong motivator. People work for it, go to great lengths to achieve it, and guard it often at great expense. Thus, it is very likely that brain

regions that respond to pleasure or rewards use these responses to decide on actions that can increase those rewards in future, or suggest actions that can explore and discover avenues for achieving those rewards.

These intuitions began to be confirmed by recordings from VTA neurons taken by Wolfram Schultz and colleagues. In these experiments, which were conducted in three stages, electrodes were inserted in the VTA of a monkey, and the animals, in this case monkeys, were allowed to reach out to a food object, a piece of apple, hidden inside a box. When the animal touched the piece of apple, dopamine neurons starting firing away at a higher than normal frequency as expected. Thus, by direct recording from dopamine neurons, and not more indirectly by functional imaging, it was confirmed that dopamine neurons respond to food rewards. In order to confirm that the stimulus that elicits dopamine cell responses is the food object, and not something else, the experimenters kept the box empty on a few occasions. When the monkey's hand touched the bare wire in the middle of box, without the piece of apple, there is no dopamine cell response.

In the second stage of experimentation, the experimenters paired the presentation of food with a neutral stimulus like the ringing of a bell. A bell is first rung, and then, after a delay, the animal is allowed to grab the food in the box. Thus, the ringing of the bell is *predictive* of the opportunity to get the reward. This time the dopamine neurons showed a briefly heightened firing rate right at the time when the bell is rung, but there was no change in firing rate when the food was obtained. Thus, it appeared that the firing of dopamine neurons represents not actual rewards but *future rewards* that the animal is expecting to obtain.

The experiment was slightly altered in the third stage. The bell was rung and the dopamine neurons briefly increased their firing rate as before, but at the time of presentation of food, the experimenter cheated the animal and did not place the food in the box. Therefore, when the animal extended its hand to reach out for the fruit, it found the box empty. At this time, there was a brief reduction in the firing rate of VTA dopamine neurons. It is as though this brief fall in firing rate represents the "disappointment" that the animal might have experienced when the food reward that it was expecting to arrive at certain instant did not occur, or when its expectations did not match with the reality. Thus, the third stage of the experiment suggested that the firing of dopamine neurons indicates not the present or future reward but the discrepancy between the expected future reward and the actual future reward.

These findings gave an important clue regarding what the brains might be doing with dopamine cell responses to rewards. Imagine an experimental animal that is permitted to press one of two buttons—A and B. Pressing button A fetches a reward (say, a piece of apple), whereas when button B is pressed nothing happens. When the animal presses button A, dopamine neurons increase their firing. This signal enables the animal to learn to choose A over B so as to continue to get more reward. Thus, the dopamine signal helps the animal to choose rewarding actions over unrewarding ones.

Although it is a dramatic simplification, choosing rewarding options over unrewarding ones is what decision-making is all about. Whether the decisions refer to larger problems of human life (what job? which partner? etc.) or the simpler ones of

an experimental animal (which button?), decision-making is essentially about choosing rewarding actions. This ability by which an organism learns to map stimuli to actions with the help of feedback from the environment in the form of rewards (or their opposite—punishments) is known as *reinforcement learning*. A lot of animal and human behavior can be explained using concepts of reinforcement learning.

Thus, the purpose of the brain's reward system is not just to create a sense of *joie de vivre* but something more fundamental, essential to the survival of the organism, namely, decision-making. The reward system is a sort of a navigational system enabling the organism course through the labyrinthine paths of life, taking rewarding turns and avoiding punitive ones.

We are now left with an important question for which unfortunately there is no easy answer. What is the relationship between the fear or punishment system, that we encountered in the earlier parts of this chapter, and the reward system just described? The reward and punishment system form the yin and yang of the brains emotional network, interacting and informing each other. But it is difficult to precisely delineate anatomical boundaries to these two systems for several reasons. Dopamine neurons which are generally considered to respond to rewards are also found to be responding to aversive or punitive stimuli. Similarly, the amygdala, which has been introduced earlier in this chapter as a substrate for fear conditioning, was also associated with reward signaling. A comprehensive understanding of brain's emotional network, with a precise mapping of each anatomical substrate to reward or punishment processing, does not exist as yet. With all the subtlety and elusiveness that is characteristic of emotions, it may be several decades before emotion researchers arrive at such a comprehensive understanding.

## Summary

We have presented an outline of how our engines of emotions work. Brain's emotion circuits are located somewhere in the middle, in the limbo between neocortex that is the stage of sensory-motor experiences, our cognitions, and other higher functions, and the low lying areas of the brain stem and spinal cord where there are centers that control our autonomic function. When we have an emotional experience, a part of the sensory stream that climbs toward the cortex bifurcates at the level of thalamus and finds its way into the emotion hubs like hypothalamus or amygdala. Activation of these centers produces two radiating influences one traveling downward and another climbing upward. The downward stream produces a wide array of autonomic responses which add to the intensity of emotional experience. The upward stream enters the cognitive, conscious world of the neocortex through the cortical window of cingulate cortex and create the emotional experience, or rather color the ongoing cognitive, sensory experience with the intensity of emotions. Thus, the element that strongly emerges in emotional experience is the connection between the higher cortical experience and the body, a connection that is established, powerfully with the densely connected hubs of the emotion circuits. The connection with the body

is more easily understood in case of animals, where the function of emotion circuits is related to primitive operations like fleeing a predator, or foraging for food. These operations obviously have a meaning to the entire organism and therefore involve a significant part of the brain and appropriate activation of the internal organs. It appears that these primitive functions of the nervous system, in their more sublime action in us, are experienced as emotions and feelings. Sensory experience is primarily limited to the sensory areas. Cognitive function engages a larger spectrum of areas, like the association areas of the posterior cortex, and the prefrontal area, in addition to the relevant sensory areas. But an emotional experience, in a sense, is not only a whole brain experience but, with its effects on circulatory, endocrine, gastroenteric, and other internal systems, evolves to be a whole body experience.

But the story of neurobiology of human emotions is far from being complete. A lot of data about emotion circuits has come from animal studies and it is nontrivial to extend these findings to make sense of human emotions. There is still quite a distance between emotions as they are understood by neurobiologists and emotions as they are depicted in the jargon of psychologists. Emotions in neurobiology are of a more primitive kind—fear, rage, satiety, pleasure, and so on, particularly in the forms that are quantifiable, measurable. But more sophisticated emotions like guilt, resentment, or gloating, emotions of the kind that show up on the outer rim of Plutchik's wheels, have not yet found their rightful place in the ganglia and goo of the real, living brain. How the primary emotional colors of fear and pleasure are transformed into the rich rainbow hues of higher human emotions is a puzzle that emotion researchers will be grappling with for a considerable time in the future. Perhaps part of the problem lies in the manner in which we seek a solution. Our approach which tries to give a name to every subtle shade of emotion, and look for neural substrates to that label, is probably fundamentally flawed. Perhaps emotions are fundamentally nonlinguistic, and therefore any attempt to neatly segregate them into clean verbal categories is probably foredoomed. Until a comprehensive neurobiological theory of higher emotions emerges on the scene, these speculations are all that we are left with. But before we give up on emotions with the argument that they are nonlinguistic, we must first consider the linguistic aspects of the brain, and describe how brain wields the power of language, a power that forms the basis for our proud position on the ladder of evolution.

## References

- Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, 32, 537–551.
- Alvarez, R. P., Biggs, A., Chen, G., Pine, D. S., & Grillon, C. (2008). Contextual fear conditioning in humans: Cortical-hippocampal and amygdala contributions. *Journal of Neuroscience*, 28(24), 6211–6219.
- Arnold, M. B. (1960). *Emotion and personality*. New York: Columbia University Press.

- Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *Journal of Neurophysiology*, 94, 327–337.
- Cannon, W. B. (1927). The James-Lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology*, 39, 10–124.
- Charles, D., Paul, E., & Philip, R. (1998). *The expression of the emotions in man and animals*. Oxford: Oxford University Press.
- Dixon, N. F. (1971). *Subliminal perception: The nature of a controversy*. New York: McGraw-Hill.
- Eustache, F., Desgranges, B., & Messerli, P. (1996). Edouard Claparède and human memory. *Revue Neurologique*, 152(10), 602–610.
- Finger, S. (1994). *Origins of neuroscience: A history of explorations into brain function*. Oxford: Oxford University Press.
- James, W. (1884). What is an emotion? *Mind*, 9, 188–205.
- Jankowiak, W. R., & Fischer, E. F. (1992). A cross-cultural perspective on romantic love. *Ethnology*, 31, 149–155.
- Kringelbach, M. L., & Berridge, K. C. (2009). Towards a functional neuroanatomy of pleasure and happiness. *Trends in Cognitive Sciences*, 13, 479–487.
- Lazarus, R. (2006). *Stress and emotion: A new synthesis*. Berlin: Springer.
- Lazarus, R., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer Pub. Co.
- LeDoux, J. (1999). *Emotional brain*. New York: Phoenix.
- LeDoux, J. (2002). Emotion, memory and the brain. In The hidden mind (special issue). *Scientific American*.
- Mobbs, D., Greicius, M. D., Abdel-Azim, E., Menon, V., & Reiss, A. L. (2003). Humor modulates the mesolimbic reward centers. *Neuron*, 40, 1041–1048.
- Olds, J. (1956). Pleasure centers in the brain. *Scientific American*, 195, 105–116.
- Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of the septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, 47, 419–427.
- Packard, V. (1961). *The hidden persuaders* (p. 41, 93). London: Penguin (Paperback edition).
- Pande, A. (1996). *A historical and cultural study of the Natyashastra of Bharata* (p. 313). Jodhpur: Kusumanjali Prakashan.
- Papez, J. W. (1937). A proposed mechanism of emotion. *Journal of Neuropsychiatry and Clinical Neurosciences*, 7(1), 103–112 (1995 Winter).
- Prabhavananda, S. (1971). *Narada's way of divine love* (Narada Bhakti Sutras). Madras: Sri Ramakrishna Math. ISBN 81-7120-506-2.
- Robert, P. (1980). *Emotion: Theory, research, and experience: Theories of emotion* (Vol. 1). New York: Academic.
- Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69, 379–399.
- Schmitter, A. M. (2010). 17th and 18th century theories of emotions. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2010 Edition). URL: <http://plato.stanford.edu/archives/win2010/entries/emotions-17th18th/>.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80, 1–27.
- Smith, C. A. & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48(4), 813–838.
- Stoerig, P., & Cowey, A. (1997). Blindsight in man and monkey. *Brain*, 120(3), 535–559.
- Tuske, J. (2011). The concept of emotion in classical Indian philosophy. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2011 Edition). URL: <http://plato.stanford.edu/archives/spr2011/entries/concept-emotion-india/>.
- Wise, R. A. (2006). Role of brain dopamine in food reward and reinforcement. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 361, 1149–1158.

# Chapter 11

## A Gossamer of Words



*All speech and action comes readily prepared out of eternal Silence.*

—Sri Aurobindo.

It is customary in a book on brain to defer a discussion of language and its substrates in the brain to one of the later chapters. But interestingly modern neurology began in 1861, when the noted French neurologist Paul Broca discovered that damage to a certain part of the brain is accompanied with impairment in speech. It is perhaps the first instance in the history of neurology when the link between a mental function and its corresponding “seat” in the brain is clearly established. Broca’s work actually began as a reaction to the wild claims of Franz Gall, the founder of a pseudoscience known as phrenology. Phrenologists claimed that a person’s character can be read off the bumps on the head (see Chap. 1). Phrenology was the first attempt at a localization theory of brain function. But Pierre Flourens debunked many of the claims of phrenologists through his lesion studies with experimental animals. The key weakness of phrenology was its absence of experimental support. Therefore, Jean-Baptiste Bouillaud, one of the students of Franz Gall, toned down the story a bit, withdrew most of the claims of phrenology, except one. He insisted that damage to frontal lobe will be accompanied invariably with speech disorders, and challenged people to find evidence to the contrary. Broca took the challenge and began a search for patients whose profile contradicted Bouillaud’s claims.

In 1861, Broca heard of the curious case of a patient called “Tan.” Tan’s original name was Leborgne but was called Tan because that’s the only word he could utter. When Broca first met Tan, he was 50 years old and had his speech impairment already for 21 years. Broca questioned him at length about his disability. Tan replied with the monosyllable, tan, repeated twice, but accompanied by a good many gestures that were intended to be a futile compensation for his agonizing paucity of speech. Through his gestures, he could show evidence that he understood what others said. For example, he could convey quantitative information using his fingers. But the

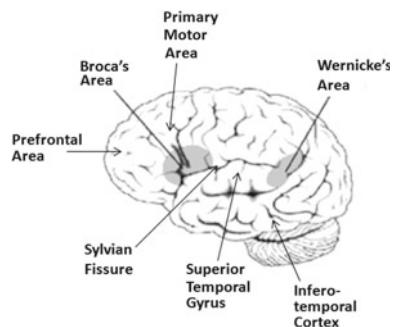
inability of his audience to understand what he said infuriated him, and he expressed his displeasure using a highly constrained form of swearing—repeating “tan” twice in quick succession! After about 10 years of this difficulty, a new problem began to set in. The muscles of his right arm gradually weakened, a weakness that culminated in complete paralysis. Subsequently, the damage spread to the right leg too, which confined the patient to the bed. After many miserable years of near immobility, Tan died in 1861. A postmortem study of his brain revealed a large lesion in his left frontal lobe.

Over the next 2 years, Broca discovered 12 more cases of similar speech impairment. One of them was a person named Lelong who could only say five words: “yes,” “no,” “three,” “always,” and “lelo,” the last of which is actually a mispronunciation of his name. In all these cases too, postmortem studies revealed damage in the left frontal lobe, specifically in a small region near the sylvian fissure. Thus, the claim of Bouillaud that the “seat of speech” is located in the frontal lobe, or, specifically, in the *left* frontal lobe, turned out to be true. This aphasia, or language disorder, in which the patient could understand speech but is severely constrained in producing it, has come to be known as Broca’s aphasia. Aphasia refers to a language disorder that affects all modalities of language comprehension or expression—reading, writing, listening, and speaking. Broca’s aphasics could utter a few isolated words but are incapable of forming long complex sentences.

The existence of a specific brain area that is responsible for speech production seemed to confirm the localization view of brain function. Subsequent discovery of a separate area for language comprehension seemed to further strengthen this view. Carl Wernicke a German physician who lived in the second half of the nineteenth century discovered that damage to the superior temporal gyrus, mostly in the left hemisphere, resulting in inability to understand language, spoken or written. Wernicke called this area the “area of word images” and proposed that it is the site of language understanding. Patients suffering from Wernicke’s aphasia, as this particular type of aphasia is called, could reflexively repeat, parrot-like, speech sounds that they hear. But they do not show any evidence of understanding what they heard. Since they do not understand what they hear, what they say too is not quite intelligible. For example, a line from a Wernicke’s aphasic: “I called my mother on the television and did not understand the door. It was too breakfast, but they came from far to near. My mother is not too old for me to be young.” The sentences are grammatically well formed but the words are inappropriate, and wrongly substituted (“television” for “telephone” and “breakfast” for “fast”). Postmortems of this class of aphasics revealed a lesion in the posterior part of the left superior temporal gyrus. Wernicke proposed that this eponymous region (Fig. 11.1) is responsible for language comprehension.

With the discovery of separate sites for language production and language understanding, the case for localization grew stronger. Encouraged by his success, Wernicke predicted the third form of aphasia, known as conduction aphasia, which should arise due to damage to a hypothetical pathway that connects the Broca’s to the Wernicke’s area. Wernicke guessed that such a pathway must exist, since it would be essential to repeat what you heard. Such an aphasia was indeed discovered subse-

**Fig. 11.1** Broca's and Wernicke's areas in the brain



quently and, as expected, patients suffering from conduction aphasia were unable to repeat what they heard. These patients had lesions in arcuate fasciculus, a pathway that was found to connect Wernicke's area to Broca's area.

Wernicke could now see the simple localization view giving way to a more satisfactory synthesis of the localization and the aggregate field views. Both views are true in their own way, and it is important to understand how they can be reconciled. If we consider language function as a case under study, language production occurs in Broca's area, while language comprehension is accomplished in Wernicke's area. In that sense localization is true. But these two areas do not work in isolation; they are connected and form a network. Therefore, language processing at large is conducted by interaction between Broca's and Wernicke's areas, and possibly other areas that handle the relevant sensory function (reading text or listening to speech sounds) and motor function (controlling the muscles involved in speaking, and gesturing). We thus have an outline, a first approximation, of the architecture of language processing in human brain.

Wernicke's synthesis is admirable for its conceptual beauty, for its neat packaging of receptive and expressive language functions in separate brain areas, and their interactions subserved by a physical wiring system. Such a picture would have served a great purpose in Wernicke's time to resolve the long-standing debate over local versus global views of brain function, in shifting the attention to networks from atomistic "seats" of brain functions. But its merits probably end there. For Wernicke's tidy conceptual picture was not fully corroborated by subsequent experimental data.

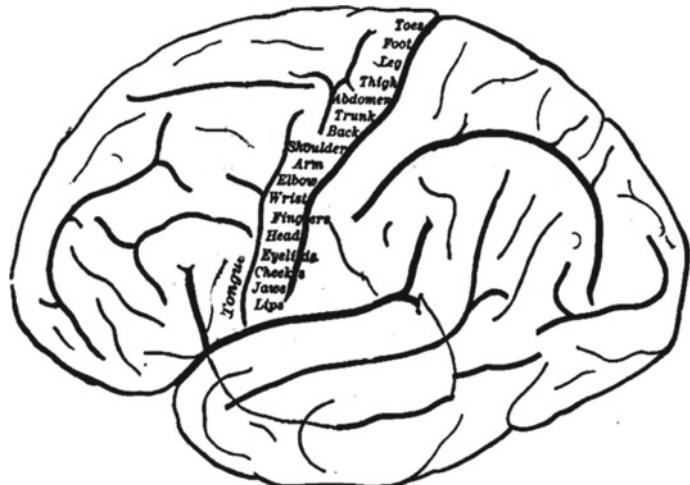
For example, Broca's area is not a pure language production center as it was originally made out to be and there is evidence that points to its role in language comprehension. Broca's aphasics who have difficulty in producing complex grammatically correct sentences also showed a corresponding difficulty in understanding the syntactic structure of sentences. Imaging studies showed activation of Broca's area also when the subjects listened to and comprehended complex sentences. Thus, Broca's aphasia is not a pure expressive aphasia.

Furthermore, since Broca's early work, there has been a growing awareness that Broca's area is not crucial for speech production. A few years ago, long after Broca's work, the brain of his two patients ("Tan" and Lelong) was studied using MRI. A key

finding of this study was that the area that Broca identified as the one responsible for Broca's aphasia is not the same as the area that is now recognized as Broca's area. Furthermore, a lesion that is exclusively confined to Broca's area probably causes temporary disruption of speech but not a permanent cessation of speech. Therefore, it appeared that lesions in areas other than Broca's area would have contributed to the severity of speech impairment in Broca's patients. For example, in one patient with a glioma (tumor of the glial cells), surgical removal of the tumor destroyed left inferior and middle frontal gyrus (cortical areas that include Broca's area) and other subcortical areas. Soon after the surgery, the patient exhibited difficulties in understanding complex sentences involving more than two subjects. The patient also had difficulty in producing reported speech. Some of these impairments were later attributed to difficulties related to working memory and not directly to speech production in Broca's area. But these problems were minor and the patient was able to resume his professional activity as a computer engineer 3 months after the surgery. Two take-home lessons emerge from these findings. First, Broca's area is only one area—probably an important one—among a network of areas responsible for speech production. Second, due to the rich compensatory mechanisms that have their roots in neural plasticity, brain can reorganize itself in such a way that speech production at large is minimally impaired in spite of a focal damage to Broca's area.

Similarly, the neat definition of Wernicke's area as a language comprehension area also turned out to be fuzzy. Wernicke's aphasia, which is a problem of language comprehension, is also accompanied by difficulty in language expression. These aphasics use words in wrong contexts, produce wrong words, omit words, or sometimes express a cascade of meaningless word strings, a phenomenon known as "word salad." Thus, Wernicke's aphasia is not a pure receptive aphasia. There is another difficulty with the idea of a "center for language comprehension" since there is really no such singular entity. There is a lot of debate about the exact anatomical demarcation of Wernicke's area. According to the classical definition, Wernicke's area is located in the posterior part of the superior temporal gyrus. But others have located it in the auditory association area, a part of the cortex that extracts higher level concepts from auditory information, also located in the superior temporal gyrus but anterior to the primary auditory cortex. Other definitions of Wernicke's area have included multimodal association cortical area in the temporal lobe. This area corresponds to the highest levels of auditory, somatosensory and visual processing streams. Higher level concepts buried in these sensory modalities, but transcend these modalities, are identified in this area. The word "rose," for example, whether heard or seen in printed form, might produce equal response in the association area. Thus, Wernicke's area is not a single clearly defined anatomical region, but a loose network of nearby regions, tied to the uppermost levels of sensory processing hierarchies.

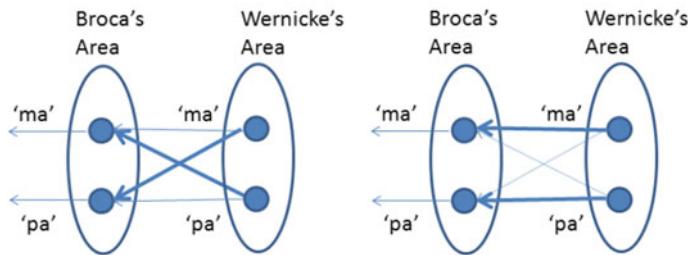
From the above considerations, it is clear that a sharp polarity between the Broca's area as speech production area, and the Wernicke's area as a language comprehension area, is an invalid argument. The strong interdependencies between these two areas can be rationalized if we probe slightly deeper into the origins of these two areas. Note that a speech production area is primarily a motor area. One of the strong organizational principles of the brain, particularly of the cortex is that areas with



**Fig. 11.2** Organization of primary motor cortex, M1. Speech-related organs like lips and jaws are controlled from inferior areas, while legs and toes are controlled from dorsal areas of M1

similar function are close to each other. Therefore, it is only natural that the Broca's area is close to the motor cortical areas. Particularly, it must be noted that the motor homunculus is organized such that muscles of mouth and lower jaw are controlled from inferior regions of the primary motor area, while arms and hands are controlled from more dorsal areas (Fig. 11.2). Therefore, it makes natural sense that Broca's area, which controls the organs of speech, is also located close to the inferior parts of the primary motor cortex. Similarly, language comprehension involves processing auditory information and extracting language content from the same. Therefore, it is natural to find Wernicke's area, or the several areas that have been identified with Wernicke's area, close to or within the sensory association area.

The above line of reasoning explains one more key feature of the anatomical location of Broca's and Wernicke's areas. Why are these two areas, belonging, respectively, to very different (sensory vs. motor) domains of brain function, close to each other? Note that Broca's area is located in the perisylvian ("about the sylvian fissure") area of the frontal lobe, while the Wernicke's area is located in and around the superior temporal gyrus, which forms the inferior bank of the sylvian fissure. It is likely that this adjacency is also driven by the general topographic organizational principle of the brain. When a part of the Broca's area is activated, and a speech sound is uttered, the sounds that are heard activate corresponding parts of the Wernicke's area. Thus, speech uttered produces correlated activity between corresponding parts of Broca's and Wernicke's area. This repeatedly occurring correlated activity between a part of the motor area, and a part of the sensory association area might be responsible in bringing these two areas to the nearest possible locations—the two banks of the sylvian fissure. In fact, the presence of correlated activity in corresponding neuronal pools of Broca's or Wernicke's areas could be playing an important role



**Fig. 11.3** A simple schematic that shows how Hebbian learning can shape mappings between Wernicke's and Broca's areas

in the development of language acquisition. According to the so-called babbling hypothesis, during the infant's first year, repeated articulations of certain syllables could induce correlated activity in corresponding pools of neurons in Broca's and Wernicke's areas. Hebbian-like plasticity mechanisms could strengthen the connections between the corresponding neuronal pools, thereby creating a strong functional coupling between the two areas.

Consider a small hypothetical network, representing an infant's language network, in which the tiny Wernicke's area projects to the Broca's area (Fig. 11.3). The Broca's area has only two units, which when activated can produce two syllables, "ma" and "pa," respectively. The Wernicke's area too has only two units, which are activated when the syllables "ma" and "pa" are heard, respectively. Although both units of Wernicke's area are connected to both units of Broca's area, the strengths are wrong to start with (Fig. 11.3a). For example, the "pa" unit of the Wernicke's area has stronger connection to the "ma" unit of Broca's area, and so on. Now let us allow the network to babble randomly, and adjust the connections according to Hebbian plasticity. When the network utters "pa," the "pa" units of Broca's and Wernicke's areas are activated in that order. Correlated activation of these units increases the strength of the connection between the units which was originally weak. Furthermore, since the "pa" unit of Broca's area and "ma" unit of Wernicke's area do not show correlated activity, the connection between them weakens. After adequate training of this sort, the connections between the two areas are likely to evolve to the configuration shown in Fig. 11.3b.

Therefore, it is clear that the path from the Wernicke's to the Broca's area is not an open branch, but forms a closed loop. What is uttered is also heard and reinforces the associations between word forms and their articulation. Once the two language areas form a loop, it is easy to recognize how inextricably linked the functions of the two areas are.

## Ascent of the Word

We began with a simplistic picture of language circuit in the brain: there is a language comprehension area connected to a language production area. Soon we pointed out the flaws in this picture and showed that there is a lot of sharing of function between these two areas. The two areas too are not unitary entities but are complex networks of nearby areas on either side of the sylvian fissure. But such a description still presents language processing in the brain as if it occurs in isolation, independent of all else in the brain, and therefore makes it somewhat mysterious. The first step toward demystification of language processing in the brain is to consider it as a form of sensory-motor processing. Language inputs are presented to the brain in visual (printed text), auditory (spoken word), or even tactile (embossed Braille-like text) forms, and the output is in the form of speech, handwriting, or even signing (as in a sign language), all of which involve activation of specific sets of muscles. Therefore, language processing is a special case of sensory-motor processing.

A hint of this link between sensory-motor processing and language processing was already given in the previous section in an attempt to rationalize the anatomical locations of Broca's and Wernicke's areas. Let us pick up this thread of viewing sensory-motor processing as the substratum of language processing and look further. How does the more general sensory-motor processing get channelized into the more specific form of language processing? How does a word, presented in whatever sensory modality, ascend through the corresponding sensory hierarchy and find its way to the language circuit? How are these two systems—the more general sensory-motor system and the specialized language system—related?

Neuroscientists Antonio Damasio and Hanna Damasio suggest that there are not two but three brain systems that support language processing in the brain. The first is the general sensory-motor system located in both the hemispheres, supporting a variety of interactions between the body and the environment. These sensory-motor images when categorized give rise to concepts. The second is a smaller system, usually located in the left hemisphere, representing the components of language, the entire hierarchy of phonemes, syllables, and words, and the rules of syntax that specify the manner in which these components can be arranged together. The third system, particularly emphasized by the Damasios, serves as a two-way bridge between the other two systems. This system takes the sensory-motor image of a concept in the first system and passes it on to the second system, enabling activation of specific word forms, networks of neurons that respond to specific words. Or when a word form is activated in the second system, the third system passes it on to the first system enabling activation of the corresponding sensory-motor image, or the concept associated with the word. For ease of reference, we will call the above three systems (1) the concept system, (2) the language system, and (3) the bridge system, respectively.

Let us visualize how the above three systems work together in handling a simple concept—"running." The act of running involves rhythmic activation of the muscles of lower and also upper extremities. It is also accompanied by the feeling of rising body heat, of the exertion, and the exhilaration if the running happens to be a part

of a race, and one is on the verge of winning. This whole sensory-motor experience associated with running supplies the raw material for the concept of running in the concept system. When we look for a word to convey this concept, essentials of this concept are conveyed, via the bridge system, to the language system, and a web of neurons corresponding to the word “running” are activated. The language system may choose to express this word through speech, or handwriting or other means.

Antonio Damasio and Hanna Damasio offer color processing and color naming as an example of this tripartite organization of language circuits in the brain. We are able to see the world in color because different photoreceptors in the retina of the eye respond to different frequencies of light. This diversity in response is preserved all the way to the visual cortical areas in the occipital lobe, e.g., the primary visual area, known as V1, and two higher visual areas known as V2 and V4. Among these areas, V4 is particularly involved in processing color. Damage to V4 causes a condition known as achromatopsia, or inability to perceive or experience color. The world appears to achromatopsics like a classic black-and-white movie, all in shades of gray. Thus, as far as color perception is concerned, V4 may be treated as the concept system.

If the lesion is near the Wernicke’s area, in the left posterior temporal or inferior parietal areas, not too far from V4, color concept is preserved, but the patient has difficulty in accessing the names of colors. They show evidence that they understand the color, experience the color. They can match and segregate colors, but may fail to retrieve the relevant word form accurately. For example, a “yellow” might turn out to be a “hello.” The vicinity of Wernicke’s area, therefore, corresponds to the language system.

Damage to the third area causes difficulty in connecting the color names with color concepts. This area is located in the temporal part of lingual gyrus of the left hemisphere. The lingual gyrus is located on the medial, or inner, surface of the brain and is not visible from outside. It stretches from the occipital pole in the posterior end and, passing via the temporal region, it extends all the way to the parahippocampal gyrus. This name “lingual” does not refer to the putative role of this area in language, but to its tongue-like shape. Patients with lesion in this area have difficulty in matching color names with color concepts. They had access to color names independently, and they showed evidence of understanding color (color matching, grading hues, etc.). For example, they might be able to match the color of a cricket ball with that of a tomato. But when queried about the color of the ball, they might respond with a “blue” or “green.”

But the clear-cut tripartite organization has been seen in color processing and naming need not be shared by other concept categories. The exact site of concept processing also varies depending on the category of the concept. Unlike in the case of language comprehension and production, there is no single brain area that is responsible for processing concepts underlying words, or the meaning of the word. In fact, the manner in which the meaning of a word is represented in the brain is a key challenge in contemporary neuroscience of language. A strict tripartite organization into language, concept and bridge systems may not be possible in every case. But one thing is certain. There is a language system consisting of the Wernicke’s and Broca’s

areas, usually in the left hemisphere, at the core, supported by a more widespread concept or word meaning system, spread over both the hemispheres.

Our quest to understand how words are represented in the brain must begin with an understanding of what a word is. How does brain distinguish a word from a nonword? Does brain respond differently to word sounds as opposed to nonword sounds? Preferential treatment of word sounds as opposed to nonword sounds might begin during infancy, particularly around the sixth month when the infant begins to babble. These basic monosyllabic sounds are processed by auditory areas forming strong associations between corresponding motor (controlling articulatory apparatus) and auditory areas. These associations, which resonate strongly to self-generated word sounds, are further reinforced by the infant's early ability to imitate sounds uttered by others. Thus, the auditory–articulatory motor pathway, which is a primitive version of Wernicke–Broca's axis, evolves preferential response to word sounds to nonword sounds. Self-generated sounds are almost always word sounds, which is not the case with external sounds.

Evidence for this preference to word sounds was revealed by Magnetoencephalography (MEG) recordings on adult subjects. MEG is a highly sensitive functional neuroimaging technique that measures neural activity by detecting the extremely tiny magnetic fields that are generated by neural electrical activity. MEG recordings from the perisylvian areas of the brain showed significantly stronger responses to word sounds, compared to nonword sounds. The difference is particularly significant in higher frequencies, at about 30 Hz, which lies in the gamma range (25–100 Hz). Gamma frequencies are thought to correspond to higher cortical activity, with a putative role in conscious experience. The source of stronger response to word sounds was traced to superior temporal lobe. Thus, preferential processing of word sounds by perisylvian areas is once again confirmed.

Further specificity in brain responses was seen when different categories of words are presented. Words can sometimes be classified in terms of the manner in which the object denoted by the words is experienced. Since at a fundamental level all our experience is, or has its roots in, sensory-motor experience, words can also be classified as sensory-dominant or motor-dominant. Specifically, researchers have considered words that have predominantly visual connotation, as opposed to words with motor or action-related associations. For example, the object denoted by a word like a “giraffe” or a “koala bear” is seen in books or in movies and therefore typically experienced in visual form. But again, the case of a “cat” or a “dog” is different since, though they are also animals, it is possible to interact with them physically, as pets, and experience them in nonvisual modalities. On the other hand, a word like a “hammer” or a “wrench” has a strong motor connotation, because they are known more in terms of use than in terms of their visual appearance. In a task in which subjects were asked to silently name tools, premotor cortex and middle temporal gyrus were activated. On the other hand, when animals were to be named, occipital areas and the inferior temporal area were activated. Whereas premotor cortex is involved in motor control based on sensory (predominantly visual) feedback, inferotemporal area is involved in recognizing complex visual patterns. The middle temporal area is involved in processing moving visual stimuli. Its activation in case of tool naming probably

pertains to the moving visual patterns of tools in course of their use. A natural extension of the above line of work would be to look for differential activation of the brain in response to nouns and verbs. One study found a stronger cortical response, again in gamma range (30 Hz), close to the motor cortex for verbs, and a stronger response in the visual areas of the occipital lobe for nouns with a strong visual component. By contrast, no such differential activation was observed when action verbs were compared with nouns with a strong action-related association. However, nouns with both visual and action-related connotation showed greater responses than nouns with exclusive visual or action-related significance. Thus, the distribution of words on cortical surface is organized not on grammatical lines but on the lines of meaning of the words, or their semantics.

Using the above considerations, it seems to be possible to explain category specific lexical impairment in case of brain damage. In a study of patients who showed partial recovery from herpes simplex encephalitis, a viral infection of the brain, the patients were able to identify inanimate objects but could not identify animate objects and foods. In another study involving a case of a massive infarction in the left hemisphere, the patient showed opposite deficits: ability to identify animals, foods, and flowers, and inability to identify certain categories of inanimate objects.

The topography of word maps in the brain goes beyond the coarse dichotomy of “action verbs in frontal area” versus “visually related nouns in occipitotemporal areas.” A study conducted by Friedemann Pulvermuller and colleagues considered three subcategories of action verbs referring to the body part with which the action is performed. The three categories refer to actions performed with the legs (e.g., running), arms (e.g., waving), and mouth (e.g., eating). When words in three categories were presented, the perisylvian areas were activated uniformly in all cases. But, in addition, the word categories produced unique responses in selective areas of primary motor cortex, which is located on the precentral gyrus along the dorsal/ventral axis (Fig. 11.2). Legs are controlled by neurons in the uppermost (dorsomedial) portion, and leg-related action verbs produced activation in this area. Arms are controlled by the central portion, which showed activation when arm-related verbs were presented. Mouth and articulator muscles are controlled by the lowermost (inferior) part of the precentral gyrus, which responded to presentation of mouth-related words. In another EEG-based study, Pulvermuller and colleagues compared brain responses to two categories of action verbs: face-related verbs (like “talking”) and leg-related verbs (like “walking”). Again, a strong activation in the inferior part of primary motor cortex was observed in response to face-related verbs, while leg-related words produced selective responses in the dorsomedial regions.

Antonio and Hanna Damasio discuss clinical studies that reflect differential brain responses to various word categories. One of their patients, known as Boswell, had difficulty recognizing specific entities—a specific place, object, or an event. He could not recognize many classes of animals, though he could detect that they are living entities. When shown a picture of a raccoon, for example, he would say that it is an animal, but could not relate to its features—its size, habitat, and life pattern. But Boswell did not have difficulty with other types of objects which have an action associated with them, like, for example, tools and utensils. He can relate to abstract

notions like beauty, ugliness, and love. He can understand actions like jumping and swimming. He can also comprehend “glue-words,” words that denote abstract relations among objects and events, like “above,” “under,” “in,” “out,” etc. He had no impairment in grammatical ability and could form syntactically correct sentences. Boswell had lesions in left temporal pole and anterior temporal cortex.

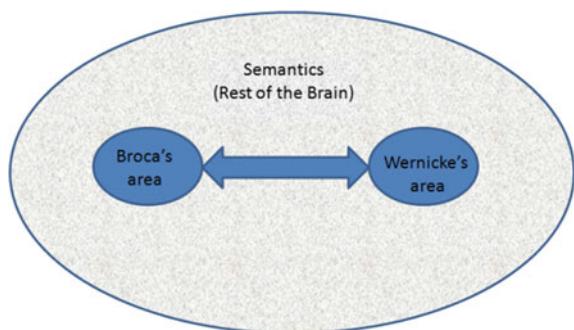
Similar patterns of word category-dependent brain activation were reported by Alex Martin and colleagues at the National Institute of Mental Health in Maryland. Using positron emission tomography, this group found that in a task involving naming pictures of animals, medial areas of left occipital lobe are activated. On the contrary, activation of left premotor cortex was observed in a task involving naming tools. In addition, left middle temporal gyrus, an area involved in processing visual motion, is activated probably due to the association with moving images of tools in action.

Thus, different researchers discovered different brain areas, other than the core perisylvian areas, responding to words, depending on the sensory-motor associations of the words. Such response patterns can, in general, be explained in terms of correlational mechanisms of Hebbian learning. Consider two neuronal pools, A and B, located inside Wernicke’s area and the visual cortex, respectively. In case of a novel object, when name-object associations are not yet formed, the sound of the word denoting the object evokes response only in area A, while the visual presentation of the object evokes response only in B. At this stage, the connections between A and B are weak or nonexistent. But when the word sound and the visual stimulus associated with the object are repeatedly presented simultaneously, both A and B are activated together. This correlated activation of A and B triggers Hebbian plasticity between the two neuronal pools. Subsequently, activation of A by the word sound automatically activates B due to the strong A-B connections, even though the object is not visually presented.

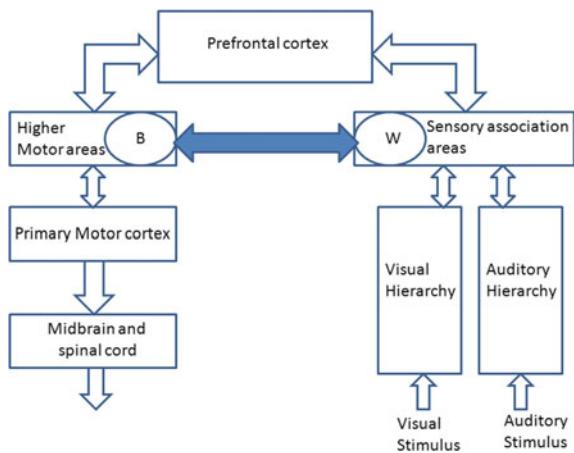
In the beginning of this section, we have set out to expand the simplistic picture of language circuit of the brain as consisting of only two areas—Wernicke’s area projecting to Broca’s area via arcuate fasciculus. We proceeded with a discussion of where semantic information, i.e., meaning of words is represented. After reviewing an array of studies on brain responses to word categories, it became clear that unlike the core language areas of Wernicke’s and Broca’s, there is no single, unique brain area where the meaning of words is represented. Different word categories are represented in different brain areas, depending on the sensory-motor associations of the words. Considering the range of areas over which brain responses to words were seen—from prefrontal to temporal pole to occipital areas—it seems that the semantic maps are spread out of the entire cortex, if not the entire brain (a discussion of the involvement of subcortical areas in language processing is omitted for simplicity). The crisp picture of language substrates of the brain that we had in the beginning of this section suddenly grew fuzzier, leaving us with something like what is shown in Fig. 11.4.

The difficulty can be resolved if we place the core language circuit (Wernicke’s area → Broca’s area) in perspective, by showing its place with respect to other key brain areas and pathways. Figure 11.5 shows a block diagram of some of the key neural highways of the brain and their destinations. The diagram only shows the broad functional hierarchies of the brain without referring to precise anatomical

**Fig. 11.4** A simplistic view of language circuit in the brain. Broca's and Wernicke's axis constitutes the core circuit, with the semantics represented over the rest of the entire brain



**Fig. 11.5** A block diagram showing the relative position of Wernicke–Broca axis with respect to the sensory and motor hierarchies of the brain

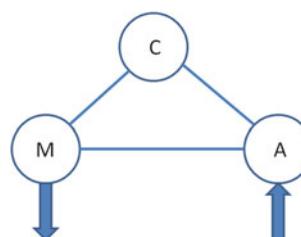


pathways and structures. At the right bottom, we see visual and auditory stimuli entering through their respective sensory channels. The two sensory streams climb through their respective hierarchies stretching across several subcortical and cortical areas. For example, the path of visual information culminating in object recognition is as follows: eyes → lateral geniculate nucleus of thalamus → primary visual cortex → secondary visual cortex → inferotemporal area. Similarly, the pathway of auditory information is as follows: ears → cochlea → inferior olivary nucleus → medial geniculate nucleus → primary auditory nucleus → higher auditory cortical areas. The highest levels of visual and auditory cortical areas project to multimodal association areas of inferior parietal cortex, where information begins to slough off its sensory coatings revealing its more abstract content. Wernicke's area, which may be regarded as a door to the core language circuit, is located at this level. There are many connections between the sensory association areas of the posterior cortex, directly with motor and premotor areas, or via higher areas in the hierarchy like the prefrontal region (Fig. 11.5). The Wernicke's area (W) → Broca's area (B) projection (shown as a filled double arrow in Fig. 11.5) is only one among the massively parallel projections from the posterior cortex to the frontal cortical areas.

Within this big picture, it now seems natural that when brain processes words of different categories, in addition to the  $W \rightarrow B$  pathway, the surrounding areas like sensory cortices, the sensory association areas in general, the prefrontal cortex, and the higher motor areas are also activated. It is now clear why it appears that word meaning is represented nearly “everywhere else” in the brain. In this larger picture, we can think of area W as another sensory association area and area B as another higher motor area. When we read aloud, or listen and respond by speaking, or perform any other sensory to motor transformations involved in language processing, this transformation is done over the hotline of W-B circuit, but also concomitantly over the more widespread word meaning or semantic webs in the brain. Thus, the W-B pathway may be regarded as a “direct route” from the input to the output side of language processing, as opposed to more indirect routes running parallel, outside the W-B pathway. The existence of such a “dual route,” as we will see shortly, has important consequences for the way we perform input–output transformations of language, under normal conditions and conditions of brain disorder.

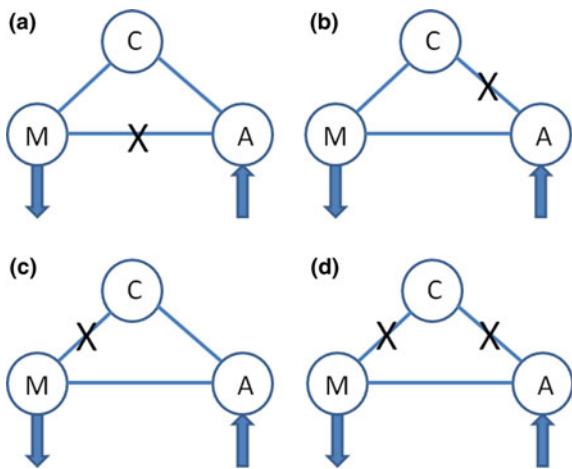
## Mechanisms of Reading Words

Shortly after Wernicke proposed his grand synthesis of local versus global views of brain function, the language circuit consisting of Wernicke’s and Broca’s area, a German physician named Ludwig Lichtheim presented an alternative view, one that expanded upon Wernicke’s view. Unlike Wernicke’s circuit which consisted of only two language centers, Lichtheim’s circuit had three centers (Fig. 11.6). Language input to the brain in the form of spoken words are presented to the “acoustic language center” (A), which is analogous to the Wernicke’s area. According to Lichtheim, A stores and recognizes images of sound sequences of words. M represents “motor language center,” which contains representations of articulatory movements made to pronounce a word. The third center, a novel one absent in Wernicke’s view, is the “center for storing concepts,” C. In modern terms, C corresponds to the extensive word meaning areas in the brain that we encountered in the previous section.



**Fig. 11.6** Lichtheim’s scheme for language processing in the brain, which consists of an articulatory center, A, analogous to the Wernicke’s area, a motor language center, M, analogous to the Broca’s area, and a concept center, C, where concepts that correspond to words are stored

**Fig. 11.7** Specific damages in the Lichtheim circuit correspond to specific forms of aphasias. **a** Conduction aphasia, **b** transcortical sensory aphasia, **c** transcortical motor aphasia, and **d** double mixed transcortical aphasia



The beauty of Lichtheim's scheme is that it was able to predict and explain a number of aphasias known at that time. For example, consider the situation when the branch directly connecting A to M is damaged (Fig. 11.7a). Conduction aphasics can comprehend speech well; they can also speak fluently, but their speech is flawed (what is known as *paraphasic*). For example, "I like to have my soap (soup) hot." But these patients fare miserably in speech repetition tasks. This inability to simply repeat what is heard has its roots in the damaged direct connection between A and M.

Now consider damage to the branch connecting A to C (Fig. 11.7b), a condition that is responsible for Transcortical Sensory Aphasia (TSA). Like conduction aphasics, patients with TSA can produce fluent but paraphasic speech. They have difficulty in thinking about and recalling meaning of words, an activity that depends on the link between the word form representations of A, with their meaning available in C. Up to this point, the impairment exhibited by TSA patients are very similar to those of Wernicke's aphasics. These patients, however, are able to repeat the words they hear, sometimes even compulsively, a phenomenon known as echolalia, thanks to the intact A-M pathway. Wernicke's aphasics do not exhibit echolalia.

Damage to the C-M branch (Fig. 11.7c) causes Transcortical Motor Aphasia (TMA). TMA patients have good language comprehension since A is intact. They can repeat simple sentences since the A-M branch is intact. But their difficulty lies in generating spontaneous speech, which involves drawing content and concepts from C. Therefore, their speech is non-fluent, halting, and effortful. Their sentences are short typically only one or two words long.

Finally, damage to both A-C and M-C branches (Fig. 11.7d) causes a more bizarre form of aphasia. Dubbed mixed transcortical aphasia, in this form of aphasia the Wernicke's–Broca's pathway is intact but the links that connect this pathway to other brain areas that participate in language processing are broken. It is as though the core language circuit is isolated from the rest of the brain. This happens because the areas

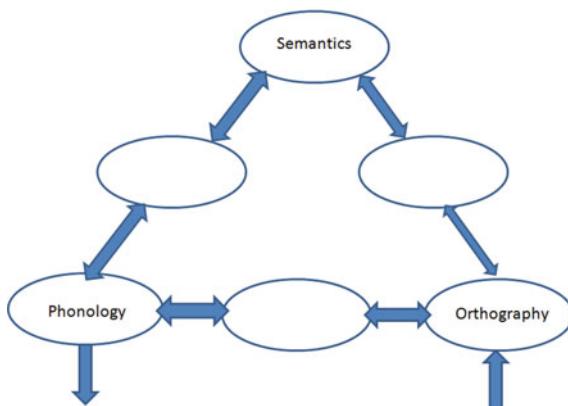
surrounding Broca's and Wernicke's areas are damaged. Transcortical aphasics have difficulty understanding speech and have poor spontaneous speech. But they can repeat complex sentences, or complete a song when the beginning is played out to them.

With its new addition of a Concept module (C), the three-module (A, M, and C) network of Fig. 11.6 has succeeded in explaining a greater range of aphasias than the simple A-M network. In spite of the apparent success, we must note that what we have so far is only a vague, high-level description that does not give sufficient insight into how the proposed modules work and interact. For example, how exactly does the Wernicke's area "process" speech sounds? And how does the Broca's area "program" the articulatory apparatus and produce speech? How exactly is the putative transformation between what is heard and what is said conducted over the fibers connecting A and M modules in Fig. 11.6? There have been attempts to make the above intuitive deliberations of how the modules A, M, and C perform, with the help of computational neural network models.

We have seen in Chap. 4 how multilayer neural network models have been applied to language-related tasks like past-tense learning and reading text aloud. Similar connectionist models applied to the problem of language processing in the brain have proven themselves to possess a great explanatory power. Connectionist models have been used to explain not only expressive disorders (aphasias) but also receptive disorders, or disorders of reading and comprehending known as dyslexias.

A general connectionist scheme for processing single words has been proposed by Mark Seidenberg and James McClelland. A slightly modified version of the same was described by David Plaut. This scheme, like the Lichtheim scheme, has three clusters of neurons, representing orthography (O), phonology (P), and semantics (S) (Fig. 11.8). Orthography denotes the graphical structure of the visually presented word, loosely translatable as "spelling." Phonology refers to the breakup of the word into its component phonemes, the atomic, basic sounds out of which the sounds of any language are constructed. Languages differ in the phonemes that constitute them, and all sounds are not phonemes. Semantics refers simply to word meaning. Orthography may be partly mapped onto the module A in Lichtheim's scheme, or Wernicke's area, though there is a key difference. Wernicke's area processes language inputs in several modalities—auditory, visual, etc.—while orthography refers to the visual word form only. The "phonology" region (Fig. 11.8), however, which refers only to language sounds, is closer in spirit to the Broca's area. The semantics area at the top of Fig. 11.8 corresponds to the extensive word meaning areas visited in the previous section. There is a hidden layer of neurons between orthography and phonology, which will enable more complex transformations between the two regions possible. Likewise, there are two hidden layers connecting semantics with phonology and orthography, respectively. Remember that a multilayer neural network, which has at least one hidden layer, has the property of universal approximation—it can learn practically any function if there are an adequate number of neurons in the hidden layer. Strictly speaking, the subnetworks linking the three modules of Fig. 11.8 are not multilayer perceptrons, because the three modules are linked in a loop and multilayer perceptrons, by definition, do not have loops. Each layer of neurons in the network

**Fig. 11.8** Schematic of a connectionist model that resembles Lichtheim's scheme. Similar to A, M, and C modules in Lichtheim scheme, this connectionist scheme consists of orthography, phonology, and semantics modules, respectively



of Fig. 11.8 is a Hopfield-type network of dynamic neurons. Information about phonology, orthography, or semantics of a given word is represented as distributed activity in the respective layers. Similar words are associated with similar patterns of activity in their respective modules. Although the subnetworks of Fig. 11.8 are not multilayer feedforward networks, the presence of hidden layers is valuable in this case too. Multilayer networks with Hopfield-type neurons and hidden layers can learn a much greater variety and number of patterns than a single-layer Hopfield network.

Words are represented as distributed activity patterns in the neuron layers in the architecture of Fig. 11.8. Similar words are represented by similar distribution patterns in the appropriate layer. For example, words with similar orthography (“race” and “lace”) have similar representations in the orthography layer. Or, a word pair like “bean” and “been” would have similar representations in both orthography and phonology. Transformations between different layers are represented in the connections between them, which are trained gradually using a procedure similar to the backpropagation algorithm described in Chap. 4.

Although Seidenberg and McClelland proposed the general conceptual scheme, they have taken a more specific problem and performed a simulation-based study. The aim of this study is to understand how people read simple words. A three-layer network with the orthography as the input layer and phonology as the output layer are chosen for this simulation. The network was trained on about 3000 monosyllabic words. The word spelling (orthography) is presented as input to the network and the pronunciation (phonology) is obtained as the output. A sample spelling–pronunciation pair would be: “gel” (spelling) and “jel” (pronunciation). The network learnt the mapping remarkably well making mistakes on less than 3% of the words. The pattern of errors produced by the network is revealing and provides insights into reading patterns of real people.

Some of the errors made by the network are what may be described as “regularization” errors. English pronunciation is plagued by a good number of exceptions to every rule. The vowel sound denoted by “oo” is pronounced differently in “book”

and “boon.” Similar is the case with the vowel “u” in “cut” and “put.” These exceptions arise because pronunciation of vowels is not consistent and is resolved by the context. Although vowels are a key source of spelling–pronunciation errors in English, exceptions can arise in pronunciation of consonants also, like, for example, the sound of “c” in “charm” and “calm.” We have addressed this difficulty in earlier in Chap. 4 in the context of past-tense learning. There is a “regular” manner of pronouncing a vowel, and there are the exceptions or the “irregulars.” Pronouncing the “irregulars” as “regulars” is called regularization. We have seen earlier that the past-tense learning network of Chap. 4 commits regularization errors in stage 2. Similar regularization errors were committed in the reading network of Seidenberg and McClelland. For example, the word “brooch” which must be pronounced as “broach” with a shortened vowel, was pronounced by the network using the regular form of “oo,” as “brUch.” Similarly, the word “spook” was pronounced as “spuk” rather than the correct “spUk.” Such regularization errors were observed in children learning to read.

In addition to errors in producing vowel sounds, there were 24 cases in which the network pronounced consonants wrongly. Some of these errors are systematic errors like use of hard “g’s” instead of soft “g’s” in “gel,” “gin,” and “gist.” But there are other cases in which the correct sound for “g” is supplied (e.g., soft “g’s” were used in “gene” and “gem”). More interestingly some of the errors produced by the network are not actually errors at all but arose due to labeling mistakes in training data. For example, the network was trained wrongly on the word “jays” as “jAs” but it pronounced the word correctly as “jAz.” This type of self-correction is possible because the network’s pronunciation of a given word depends not just on that word in isolation, but is influenced by its exposure to a large number of other similar words. The network was able to evolve a “rule” from a large number of similar words and generalize correctly to a new word, even though the word is labeled wrongly in the training set. Another class of difficult cases that the network was able to learn successfully was silent letters as in “calm” and “debt.” The network did not perform a single mistake in pronouncing silent letters. An interesting aspect of reading that emerged from the network training is the link between the network’s output error and “naming latency” which refers to the time taken by a human subject responding to the visual presentation of a word in a reading task. The response is quicker to familiar words and slower to novel and difficult words, when humans read them. Therefore, the output error (a measure of the difference between the desired output of the network and its actual output) produced by the network on a given word is likely to be related to the latency exhibited by subjects in reading. The simulation study showed that there is a simple monotonic relationship between the latency and the error. In most cases, the latencies were about 10 times the error plus a constant latency of about 500–600 ms.

Thus, the above connectionist modeling study of reading revealed that such networks can learn both regulars and irregulars with equal facility. Although there were errors, they were small in number. The success of connectionist models of reading makes a revealing statement on a long-standing paradigm that influenced earlier theories of reading. Linguists have always segregated the systematic aspects of lan-

guage (“regulars”) and the exceptions (“irregulars”). The systematic aspects were expressed as rules of grammar, spelling, pronunciation, and so on, while the exceptions are grouped separately. In the specific context of reading, there are a small set of Grapheme-Phoneme Correspondence (GPC) rules ( $G \rightarrow /g/$  or  $V \rightarrow /v/$ ) and the exceptions ( $G \rightarrow /j/$ ). This segregation into rules and exceptions has prompted the so-called dual route theories of reading, in which a separate lexical system was added to the GPCs to handle exceptions. Another approach, further down the same path, was adopted by “multiple-levels” theories, in which there were the GPCs and multiple levels of exceptions, from whole categories of exceptions to single isolated cases. Connectionist models have shown that a segregation into rules and exceptions is unnecessary; both “regulars” and “irregulars” can be represented in the same network, distributed over the network’s connections without any anatomical or structural segregation. The “rules” merely correspond to words that are more frequent, or spelling-pronunciation patterns that are more consistent. The rules are exhibited by the network as it generalizes the spelling-pronunciation patterns to novel cases.

## Understanding Dyslexia

Dyslexia in general terms means learning difficulty, with specific reference to impairment in accuracy, fluency, and comprehension in reading. Dyslexia can sometimes be modality specific, limited, for example, only to understanding visual text, or auditory inputs. About 5–10% of any population is believed to have some form of dyslexic impairment. Dyslexia could arise simply due to delayed cognitive development or due to head injury, or age-related degradation like dementia.

Dyslexia researchers defined two broad classes of dyslexia—surface and deep dyslexia. Surface dyslexics have no difficulty in reading regular words and nonwords, since nonwords are read using rules underlying regular word pronunciation. (For example, a nonword like “mave” is pronounced as “gave,” which is a standard pronunciation for “-ave,” and not as “have” which is an irregular word.) But their performance is poor on low-frequency, irregular words. In such cases, the surface dyslexics tend to regularize them. (For example, “sew” is read as “sue,” rather than the more appropriate “sow.”) The surface form of dyslexia typically arises due to damage to left temporal lobe. Anatomically speaking, surface dyslexics seem to have an intact direct pathway between orthography and phonology (Fig. 11.8), which is sometimes called the phonological pathway, with the longer route via the semantics being affected. The word meaning or semantics, imageability of the word, and other aspects help improve readability of low-frequency words with irregular pronunciation. Errors increase when this additional help, arriving via the semantic module, is absent.

Deep dyslexics exhibit impairments that are opposite to those of surface dyslexics. The phonological pathway is severely affected in these patients, which prevents them from reading even the easiest nonwords. Since the common phonological rules used for reading frequent words and nonwords are unavailable, deep dyslexics depend on

semantics even for common words. Therefore, reading errors produced by this class of dyslexics have a strong semantic component. For instance, they may read “cat” as “dog”; the words do not sound similar but have similar meaning. Or in other cases, the errors might arise in the reception of the visual word form (e.g., reading “cat” as “cot”). There are also mixed cases like reading “cat” as “rat,” in which case the two words not only have similar meaning, they also look similar.

In an attempt to understand surface dyslexia, David Plaut and colleagues simulated a network that consisted of both the phonological pathway and the longer pathway via the semantics layer (Fig. 11.8). The phonology layer receives inputs from both orthography and semantics. Conditions of surface dyslexia can apparently be reproduced by damaging the connections from semantics to phonology. Plaut and colleagues compared the performance of the “surface dyslexic” version of the network with the intact network. Performance of the damaged network was also compared with two surface dyslexic patients and the results are quite revealing. Both the damaged network and the patients showed high reading performance (90–100% correct) on four categories of words: (1) high-frequency regulars, (2) low-frequency regulars, and (3) high-frequency exceptions and (4) nonwords. But the performance was considerably low (close to 70%) on low-frequency exceptions in case of patients and the network simulation. Thus, our intuitive understanding of the dynamics of surface dyslexia is corroborated by a computational model.

A similar connection model was developed by Jeff Hinton and Tim Shallice to explain the reading errors made by deep dyslexics. The network model used in this study consisted of three layers, with the visual form of the word presented to the input layer, and the corresponding semantic information read off the output layer. The network basically transforms the visual word into its semantic representation, and the hidden layer gives to the network sufficient representational capacity. To keep the network small, and training manageable, the words with only three or four letters were used. By further restricting the letters that occur in a given position in the word, the number of words was restricted to 40. A binary representation with 28 bits is constructed to represent each of these words. Activity of the output layer neurons represents the word meaning or properties of the named object that contribute to its meaning. A list of 68 properties was considered for representing semantics. Some sample properties: “Is its size less than a foot?” “Is it a 2D object?” “Is it a mammal?” “Is it used for cooking?” “Is it a snack?” If the answer to the question is positive (negative), the corresponding neuron’s output is set to be one (zero). Another interesting feature of this network is that the output or semantic layer is modeled as an associative network with recurrent connections and attractor dynamics. Each attractor in this word denotes a word with a specific meaning. Nearby attractors in this semantic space are likely to have similar meaning. In response to a word presented at the input layer, the output layer cycles through a series of states before settling in a nearby attractor. Parts of this network were damaged and network performance on retrieving semantic information of various categories of words was analyzed. The network often misread a word as another with a similar meaning (DOG → CAT). For example, for one type of damage, the network made a significantly large number of errors on words that denote fruits. The network also made visual errors (DOG → BOG) and

confused visually similar words, since damage to the connections leading upward from the input layer is likely to fail to discriminate similar words presented to the input layer. Depending on the combination of damages implemented, the network also exhibited mixed or visual and semantic errors (CAT → RAT).

Thus, the scheme for language processing originally proposed by Wernicke, and further elaborated by Lichtheim, had enjoyed adequate neurobiological support. It established the idea of a “dual route,” which consists of two parallel systems—one “superficial” and one “deep”—for language processing. These intuitive ideas have been tested and refined by use of computational network models in both normal function and in a variety of dyslexic conditions.

We now turn our attention to a quaint aspect of language in the brain—lateralization.

## Language of the Hemispheres

In 1836, long before Broca made his discoveries, Marc Dax, an unknown small town doctor, presented a paper in a medical society meeting in Montpellier, France. Dax worked with a good number of aphasic patients and was struck by the fact that all of them had left hemispheric damage. He never found an aphasic with damage to the right hemisphere. The report which was presented to the medical society was completely ignored. Thus, a finding that could have led to one of the most exciting aspects of organization of the brain quickly ran aground.

Dax’ report was perhaps the first to observe an asymmetry in distribution of brain functions across the hemispheres. Distribution of sensory-motor function in the brain is more straightforward. Movements in the right part of the body are controlled by the motor cortical areas of the left hemisphere and vice versa. Similarly, information about the left visual field is routed by a complicated system of wiring between the eyes and the brain, to the right occipital lobe, and vice versa. A similar organization exists in case of auditory and somatosensory systems too. But language seems to be different. As Dax had first observed, we seem to understand and express language with our left brain.

About three decades later, Paul Broca also noted the interesting fact that most of his aphasic patients had left hemisphere damage. His first impression about this observation was more cautious: “...a thing most remarkable in all of these patients (is that) the lesion is on the left side. I do not attempt to draw a conclusion and I await new findings.” But by 1864, Broca was convinced that language lateralization in the brain is a real and important phenomenon.

I have been struck by the fact that in my first aphemics<sup>1</sup> the lesion always lay not only in the same part of the brain but always on the same side – the left. Since then, from many

---

<sup>1</sup>For some unknown reason, the original term that Broca used for aphasia was aphemia, which is derived from a Greek word meaning “infamous.” Noting the absurdity of this nomenclature, a critic named M. Troussseau proposed the new term aphasia, which stuck.

postmortems, the lesion was always left sided. One has also seen many aphemics alive, most of them hemiplegic, and always hemiplegic on the right side. Furthermore, one has seen at autopsy lesions on the right side in patients who had shown no aphemia. It seems from all this that the faculty of articulate language is localized in the left hemisphere, or at least that it depends chiefly upon that hemisphere.

Broca must be credited with one more observation on brain asymmetry—the link between language lateralization and handedness. Handedness is a somewhat vague idea and any attempt at a precise scientific definition brings out the complexity of the matter. But in common parlance, handedness refers to the hand that an individual uses for writing, which happens to be the right hand in a majority of people. Since the right hand is controlled by the left brain, Broca found it interesting that the side of the brain that controls speech also determines which hand is more skilled. If this correlation is a real one, Broca argued, language centers must be located in the right brain in left-handers: “One can conceive that there may be a certain number of individuals in whom the natural pre-eminence of the convolutions of the right hemisphere reverses the order of the phenomenon...” of language areas being in the left brain in right-handers.

Language lateralization proved that the hemispheres are different. But this difference soon gave way to the idea of dominance of one hemisphere over the other. Reinterpreting brain asymmetry in terms of dominance, British neurologist Hughlings Jackson wrote in 1868: “The two brains cannot be mere duplicates, if damage to one alone makes a man speechless. For these processes (of speech), of which there is none higher, there must surely be one side which is leading.” Jackson went on to conclude “... that in most people the left side of the brain is the leading side – the side of the so-called will, and that the right is the automatic side.”

Findings from patients with apraxia, an ability to perform purposeful movements on command, have reinforced the idea that the left brain is more privileged than the right one. The apraxics often have the ability to perform a movement spontaneously, but have difficulty in repeating the same on instruction. For example, an apraxic might be able to brush his/her teeth as a part of a daily routine, but not when instructed to do so. Or even when they attempt, the performance may be flawed, like, for example, attempting to brush teeth with a comb, or brushing at a painfully slow pace. Hugo Liepmann a German neurologist who did some early work with Carl Wernicke did pioneering work on apraxics. Liepmann observed that apraxia is often associated with injury to the left hemisphere. Once again left hemisphere is found to be involved in a “higher function”: it was speech in the studies of Dax and Broca, and it was purposeful or willed actions in the studies of Liepmann.

Studies of this kind have helped to shape a view that the left hemisphere is a leading or a dominant hemisphere, while the right hemisphere is a minor hemisphere without any special functions beyond the primitive sensory-motor operations. Thus, a statement of brain asymmetry degenerated into one of strong inequality of hemispheres. It led brain theorists to an absurd position according to which an entire half of the brain is bereft of any serious purpose, and condemned to the status of an unequal cerebral partner in mental life. Although there were critics of this extreme situation even in the early days of lateralization studies, their voices have been ignored. Paradoxically

one of such critics happened to be Hughlings Jackson, the neurologist who supported the “leading” role of the left hemisphere. Jackson modified his original view based on studies of a patient with a tumor in the right brain. The patient had difficulty in recognizing objects, persons, and places. Attributing a significant purpose to the right brain and restoring its position as an equal partner to the left brain, Jackson noted in 1865: “If then it should be proven by wider experience that the faculty of expression resides in one hemisphere, there is no absurdity in raising the question as to whether perception – its corresponding opposite – may be seated in the other.”

In the early 1930s, a large-scale study was conducted in order to explore the different facets of lateralization. The study which involved over 200 patients and more than 40 different tests gave remarkable results. In general left hemisphere damage resulted in poor performance on tasks demanding verbal ability. Contrarily, those with right hemisphere damage did poorly on tasks involving understanding geometric figures, solving graphical puzzles, completing missing parts of figures and patterns and so on. Right hemisphere damaged patients also showed profound disturbances in orientation and awareness of space. Some of them had difficulty in finding their way around in a house where they lived for years.

Damage to right hemisphere is also known to cause amusia, or loss of musical ability. This impairment was often observed in professional musicians who suffered right brain damage. Contrarily, left brain damage impaired speech but not musical ability. Perhaps the earliest evidence in this regard was an anecdotal record dating back to 1745. A person with left brain damage had serious impairment of speech and paralysis of the right side of the body. However, it was reported that “he can sing certain hymns, which he had learned before he became ill, as clearly and distinctly as any healthy person...”

Along with its role in musical capability, the right hemisphere also seems to have a role in prosody, which refers to stress, intonation and rhythm of speech, or in a sense the musicality of speech. Prosody is captured by properties of speech like loudness, pitch and syllable length, and therefore not captured in writing, though mildly substituted by use of punctuation and stylistic additions like italics. Emotional prosody refers to use of prosody to express emotions and feelings. Charles Darwin observes that emotional prosody dates back to an era long before the evolution of human language. In the Descent of Man, Darwin writes: “Even monkeys express strong feelings in different tones – anger and impatience by low, fear and pain by high notes.” Emotional prosody seems to be affected more in case of right brain damage than that of the left brain. Though both left and right brain damage can cause deficits of emotional prosody, the deficit seems to be redeemable in case of left brain-damaged patients but not so in case of right brain damage.

Studies and data of the kind mentioned above gave rise to a picture of brain lateralization that places language in the left brain and spatial processing in the right brain. By extension, sequential processing, of which language is an example, was placed in the left brain, and simultaneous, holistic, or pattern-thinking was placed in the right brain. Similarly, logicality which involves a sequential deduction from a premise to a conclusion was placed in the left brain. But music, an example of patterns or rhythms in time, was placed in the right brain. These oversimplifications

helped form a vague philosophy that once had its roots in science, but soon forgot them. Some have even identified Western culture with the left brain and the Eastern culture with the right brain. People were classified on the basis of the more active side of the brain.

While this kind of romanticizing of cerebral lateralization continues even today, more rigorous studies of lateralization revealed that the situation is not as simplistic as its picture painted in popular media. An experimental technique that helped collect a large amount of lateralization data on which a more concrete theory of lateralization can be erected is known as the Wada test. Using this test it became possible to temporarily anesthetize a whole hemisphere of a person's brain. In this test, a drug known as sodium amobarbital is injected into one of the carotid arteries that supply the brain with blood. There are two carotid arteries on either side of the neck, each supplying blood only to the hemisphere on which side the artery is located. Amobarbital is a GABA agonist which increases inhibitory activity in the brain, and therefore decreases general brain activation, resulting in anesthesia or loss of consciousness.

The subject participating in Wada test is asked to lie down flat on his/her back, lift his/her arms straight in the air, and count up to 100. Seconds after the drug is injected, the arm on the side opposite to the side on which the drug is injected falls limp. Furthermore, if the drug is injected on the side of the brain which controls speech, the counting stops for a couple of minutes before it is resumed as the drug effect starts wearing off. The Wada test, with its neat reversibility, and its sure shot effect on one whole side, and only on one side, of the brain made it practical to study lateralization in a large number of patients and normal subjects.

Some of the key findings of Wada test were that about 95% of right-handers have their speech and language localized to the left hemisphere. In the remaining right-handers, speech is controlled by the right hemisphere. Contrary to the general Broca's rule that left-handers have language centers on the right, it turned out that 70% of left-handers had their language controlled in the left hemisphere. About 15% of left-handers had their speech in their right hemisphere, while the remaining 15% of left-handers had speech on both sides, a situation known as bilateral speech control. Therefore, the idealized picture of language on the left versus spatial on the right is simply not true.

Therefore, it is clear that lateralization is more complex than what it is made out to be in popular media. But why is there lateralization in the first place? When sensory-motor function is present in both hemispheres, why is language processing confined to one side? It is not easy to answer the "why's" of brain's organization, since it is a product of evolution and experimental verification is not straightforward, but we can hazard a speculative explanation. It is best to begin with the reason behind lateralization of sensory function, particularly visual function. Although information regarding both the right and left visual fields enter both right and left eyes, retinal projections to the brain are organized such that the information related to the left (right) visual field arrives at the right (left) occipital lobe. Thus, a half of the visual space is seen by the contralateral brain. Other sensory streams (auditory and somatosensory) are also processed by the contralateral brain perhaps because then the same side of

the brain processes different streams of sensory information coming from a given side of the external world. Now a part of the body is controlled by the contralateral motor cortex, perhaps because in such a system, the motor cortex controlling a given side of the body is on the same side as the brain processing the sensory information from that side, thereby decreasing wire length of the connections between motor and sensory areas corresponding to the same side of the external space. This bilateral organization goes sufficiently high in the sensory-motor hierarchy. Thus, the bilaterality of sensory-motor areas has their roots in the strong bilateral symmetry of the body and bilateral organization of sensory organs—two eyes, two ears, and skin distributed nearly equally on either side of the body. But there is no reason why this bilaterality must be carried over to higher brain functions which are not closely tied to the body. Thus, language, music, prosody, understanding complex visuospatial patterns, all these high-level functions are lateralized. Lateralization of higher level functions not too closely tied to low-level sensory-motor information makes a more economic use of brain's resources and contributes to savings in wire length because bilateral organization of brain areas often requires that the corresponding areas are connected by inter-hemispherical wiring systems.

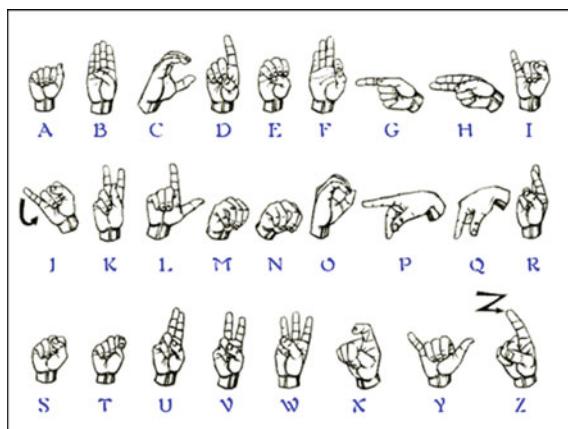
## Other “Signs” of Language Impairment

There is another form of language understanding and expression which poses a peculiar problem to functional lateralization. Individuals who have lost the power of hearing and speech communicate with each other by signing, which usually consists of gestures, hand shapes, movements of the arm, and facial expressions. Just as there are hundreds of spoken languages in the world, there are also hundreds of sign languages and there is no universal sign language. Just as its large number of spoken languages, India has a large number of sign languages used in various parts of the country. Similarly, there is an American Sign Language (ASL) which was originally derived from French Sign Language (Fig. 11.9). Even countries using the same spoken language have different sign languages. US, UK, Australia, and New Zealand have all different sign languages. So it is with Spanish speaking countries like Spain and Mexico. Wherever there was a community of deaf and dumb people, there has evolved a local system of signing.

One interesting aspect of sign languages is that, like spoken languages, they too have a lexicon and syntax. The signs are symbolic and need not have a literal visual relationship to the object that is being represented. The sign languages are complex and are capable enough to describe and discuss a range of topics from the simplest to the most abstract.

The question of neural substrates of sign languages poses an interesting challenge to neurolinguists. Considering their strong visual-spatial quality, it seems likely that sign languages are localized to the nondominant hemisphere. But since the sign languages have a sequential quality, with an elaborate grammatical structure, it is probable that they belong, like the spoken languages, to the dominant side.

**Fig. 11.9** Signs for alphabets in American sign language



This dilemma was put to test by Gregory Hickok, Ursula Bellugi, and Edward S. Klima in the 1980s, and the answer surprisingly was that sign language, like the spoken language, is localized in the dominant hemisphere. One of the earliest patients studied by Hickok, Bellugi, and Klima had incoherent signing and was seriously impaired in understanding the signing done by others. It turned out that this patient had damage in the left brain which encompassed the Wernicke's area. Another deaf patient had excellent sign language comprehension but had difficulty shaping her hands to produce signs. Her expressions were confined to isolated words and the sentences did not have any significant grammatical structure. As in the cases of expressive aphasias of spoken languages, this second patient had damage to the Broca's area.

More extensive studies were conducted beyond the preliminary ones mentioned above, which confirmed that left hemisphere damage can seriously impact sign language capability. They had difficulty identifying even isolated signs. Like slips of tongue in case of spoken languages, they committed slips of hand and substituted one sign for another. But signing ability was not damaged significantly in cases of right hemisphere damage.

But it is not that the right hemisphere has no role whatsoever in sign language comprehension and production. Right hemisphere damaged patients had no difficulty producing short, or even longer sentences, but showed deficits in maintaining a coherent discourse. Although their individual sentences are well formed, the patients rambled from one topic to another without a coherent theme linking them. Thus when it comes to sequential structure of language—even sign language—the left brain seems to be involved in an intermediate scale of the length of the narrative, but the right hemisphere seems to be necessary to create coherence at larger scales. This aspect of right brain processing is usually not brought out in aphasic studies since the test batteries used to assess aphasias consist of single sentences.

We have seen a similar local versus global distinction in the relative contributions of left and right hemispheres in visuospatial processing also. Though both left and

right hemispheres are involved in visuospatial processing at smaller scales, an intact right hemisphere seems to be necessary for comprehending larger scale visuospatial patterns. It is interesting to see if a similar local versus global dichotomy between the two hemispheres is found in performance of signers too, since signing occurs in a predominantly visual modality. Studies by Ursula Bellugi and colleagues confirmed this dichotomy in signers. When presenting a complex story, signers make extensive use of the space around them. Each character in the story is given a specific location in the ambient space of the signer. The space in front of the signer becomes a virtual stage on which the events of the story and the interactions among the characters are played out. Some signers with right brain damage could not maintain the positions of the characters on the virtual stage in a consistent fashion. Other right brain-damaged signers who had only mild visuospatial impairments in nonlinguistic tasks showed significant deficits in maintaining a coherent narrative in expressing linguistic content. Therefore, the role of right hemisphere in signing seems to lie specifically in handling large-scale visuospatial processing contributing to sign language comprehension and expression.

Thus, though it appeared that sign languages would have very different neural substrates compared to spoken languages, several decades of research has clearly established the common elements, and therefore shared cerebral substrates, between the two kinds of language at the highest levels. The apparent differences between the two forms of language are only at the lowest level of sensory modality. Spoken languages predominantly depend on auditory sources for inputs, and produce speech as outputs, whereas sign languages are visuospatial on input side and motor activity involving face and upper extremities on the output side. What is common to both forms of language on the input side is that both are sequential sources of information. Even in the case of traditional spoken languages, if we consider the written form, the input is again visuospatial, consisting of moving images in time gathered by eyes scanning the written text. In that sense, sensory processing in reading is not fundamentally different from that involved in watching signing. Therefore, just as both listening and reading access the same language processing circuit at a higher level, understanding the visuospatial content of signing also accesses the same language areas. In all these cases, at a smaller spatial scale (e.g., specific hand forms) and a smaller temporal scale (e.g., signing or saying isolated sentences), the left or dominant brain plays a crucial role in language processing. Contrarily at larger spatial scales (e.g., global organization of the space of gestures) and larger temporal scales (e.g., overall organization of a narrative, or prosody) the right or “nondominant” brain plays an important role.

We have presented an essential outline of the substrates of language processing in the brain. Irrespective of the sensory-motor modalities involved in the particular language-related operation (reading, listening, speaking, signing, etc.) we understand that at the highest levels all these modalities converge at a common circuit that is built around the Wernicke's area on the receptive side and the Broca's area on the expressive side. We have described the word meaning areas and their relationship with the core language processing pathway of Wernicke's → Broca's circuit. We have also discussed one of the most intriguing aspects of language processing in the

brain, viz., lateralization. A discussion of language in the brain paves the way to the deepest mystery, a conundrum of neuroscience that remains unresolved over the millennia: how did the brain acquire consciousness?

## References

- Backman, J., Bruck, M., Hebert, M., & Seidenberg, M. S. (1984). Acquisition and use of spelling-sound correspondences in reading. *Journal of Experimental Child Psychology*, 38, 114–133.
- Broca, P. (1865). Cited in Diamond, S. (1972). *The double brain*. London: Churchill-Livingstone.
- Critchley, M. (1970). *Aphasiology and other aspects of language*. London: Edward Arnold.
- Dalin, O. (1745). Cited in Benton, A. L., & Joynt, R. J. (1960). Early descriptions of aphasia. *Archives of Neurology*, 3, 205–222.
- Damasio, A., & Damasio, H. (1992, September). Brain and language, special issue on brain and mind. *Scientific American*.
- Darwin, C. (1871). *The descent of man and selection in relation to sex* (Vol. 1. p. 320). Murray, 1888.
- Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., & Cabanis, E. A. (2007). Paul Broca's historic cases: High resolution MR imaging of the brains of Leborgne and Lelong. *Brain*, 130, 1432–1441.
- Hickok, G., Bellugi, U., & Klima, E. S. (1998). The neural organization of language: Evidence from sign language Aphasia. *Trends in Cognitive Sciences*, 2(4), 129–136.
- Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98(1), 74–95.
- Jackson, J. H. (1958). In J. Taylor (Ed.), *Selected writings of John Hughlings Jackson*. New York: Basic Books.
- Joynt, R. J. (1964). Paul Pierre Broca: His contribution to the knowledge of aphasia. *Cortex*, 1, 206–213.
- Lichtheim, L. (1885). On aphasia. *Brain*, 7, 433–484.
- Pearce, J. M. (2009). Hugo Karl Liepmann and apraxia. *Clinical Medicine*, 9(5), 466–470.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115.
- Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5, 517–524.
- Pulvermüller, F. (2002). *The neuroscience of language: On brain circuits of word and serial order*. Cambridge: Cambridge University Press.
- Pulvermüller, F., Mohr, B., Sedat, N., Halder, B., & Rayman, J. (1996). Word class specific deficits in Wernicke's aphasia. *Neurocase*, 2, 203–212.
- Rasmussen, T., & Milner, B. (1977). The role of early left-brain injury in determining lateralization of cerebral speech functions. In S. Dimond & D. Blizzard (Eds.), *Evolution and lateralization of the brain*. New York: New York Academy of Sciences.
- Springer, S., & Deutsch, G. (1981). *Left brain, right brain*. New York: W.H. Freeman and Company.

# Chapter 12

## The Stuff that Minds Are Made of



*We do have the resources to reconstitute the body; the mind, though, will remain a gooey mess.*

*Mother character in the movie Looney tunes: Back in action.*

So far we have endeavored to demystify a few things about the brain. How is sensory information organized as maps in the brain? How are memories taken through a series of resting stages through the brain? How are fears processed and integrated with higher cognitive processes in the brain? How do simple cellular level changes underlie complex learning phenomena? Most of these processes can be shown to be implemented by networks of neurons passing neural spike signals among themselves. The circuits involved might vary, the precise nature of the signals might vary, but the broad framework remains the same and this is one of the celebrated successes of contemporary neuroscience. Using the scientific method, and the intimidating and extravagant repertoire of contemporary neuroscience methods and materials, from single neuron recordings to gene knockouts, it is possible to present a scientific theory of how brain works. This is the reason it is now possible to demystify a lot of phenomena related to the brain. But in spite of these glowing successes, there is one phenomenon that continues to be elusive and mysterious, perhaps because it is the source of everything that is mysterious. The name of that mystery is consciousness. Mystery is what mind feels when it is faced with a truth it cannot unravel. But when the mind is turned upon itself in search of the truth of its own existence, it seems to encounter a profound mystery—a mystery of all mysteries—a challenge in which a breakthrough does not seem to come by easily.

You will never know if the color blue that you experience is identical to my experience of same the color. We can have a practical, operational consensus whereby both of us point to an object we think is blue. But the experience of one remains to the other a secret, and probably remains so forever perhaps because consciousness is by its very nature absolutely personal, private. We can express contents of that

experience through words and gestures, or other representations, which are merely shadows of the real thing. This uniquely private nature of consciousness offers a serious obstacle to creation of a science of the subject, since modern science, by its very founding principles, is based on objectively sharable observations and therefore *perforce* excludes the one thing that is private by default—consciousness.

The mystery of consciousness is rendered more acute if we think of it as an attribute of a special physical object—the human brain. Why is this particular 1.3 kg of matter, of all objects in the world, endowed with consciousness? Why does not a crystal, a gas, or plasma possess consciousness? Why does not the study of any of the four fundamental forces automatically lead us to consciousness? All the ideas discussed in this book so far are comments on the complexity of this object and explain some of the key observations and capabilities of this object. But they are completely mute on how consciousness arises in this object. Furthermore, the brain is endowed with consciousness not always but at certain times and under certain conditions. A dead brain, we may safely assume is not conscious. So is a brain in coma, with very low neural activity. There are also grades of consciousness from deep sleep to dreamy sleep to light grogginess to full awakening. If human brains have consciousness, by extension must we allow animal brains, at least those belonging to the animals close to us on the evolutionary ladder, similar privileges? Primates definitely display intelligent behavior but are they conscious, in the manner we are? How far down may we go in search of the capacity of consciousness in the hierarchy of nervous systems—vertebrates, invertebrates, single-celled organisms? These questions are tossed around in contemporary philosophical and scientific debates on consciousness, with no definite answers.

The deep difficulties that crop up when the question of consciousness is brought into an objective study of the physical world were known to philosophers for centuries. The question of consciousness or mind and its relation to the physical world, or more specifically, the body, is brought into stark focus in the thought of French philosopher Rene Descartes. At the center of his thought, there are two different entities—the mind and the body. Mind is a “substance” that has no spatial extension and is capable of the power of thought, while the body has a spatial extension and is incapable of thinking by itself. Thus mind and body are very different “substances.” This distinction between mind and body was expressed in his *Meditations* as

[O]n the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing [that is, a mind], and on the other hand I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it. (AT VII 78: CSM II 54)

A deep philosophical problem arose out of this Cartesian dualism. How can two substances that are so different in their nature act upon each other as our basic experience of these two entities demonstrates. This problem known as the “mind–body problem” has no definitive solution not only in Cartesian thought but also in the intervening history of science and philosophy to date. Descartes suggested that this delicate interaction between mind and body occurs at a privileged site in the brain—the

pineal gland. This proposal did not really solve the mystery because it was not clear what was so special about the pineal gland that it made possible what was thought to be nearly impossible. Another source of difficulty in Descartes' philosophical approach is the free and arbitrary use of the idea of God to explain anything that resists an easy explanation. His disciples must have picked up this distressing practice from their master. Noting the difficulties with pineal gland-based explanations of the mind–body problem, two of Descartes' disciples, Arnold Geulincx and Nicholas Malebranche, proposed that the mind–body interactions are possible because they are presided over by none less than God! Thus, a trenchant division between mind and body referred to as Cartesian dualism—and an intermittent and inconvenient introduction of God into the philosophical arena characterize Descartes' approach to the mind–body problem and the flotsam of difficulties that came in its wake.

Descartes tried to pack too many things on his platter—from God at one end to pineal gland at the other—with little success. There were others who stayed clear of God and confined themselves to a purely psychological approach to the problem of consciousness, a notable example being that of William James. James argues that we have no right to reject outright the mind–body problem, and urges that the right modalities of that interaction must be worked out. “It is … quite inconceivable that consciousness should have nothing to do with a business to which it so faithfully attends.” According to James, consciousness is first and foremost a selection agency. It selects among multiple competing choices presented to it in the form of conscious thoughts. A crucial property of conscious thoughts is wholeness or unity; they are not a mere assembly of more elementary components. “...however complex the object may be, the thought of it is one undivided whole.” The unity and wholeness of the conscious thoughts have their roots in analogous properties in the brain. The entire brain acts together, in unison, to produce conscious thoughts:

The facts of mental deafness and blindness, of auditory and optical aphasia, show us that the whole brain must act together if certain conscious thoughts are to occur. The consciousness, which is itself an integral thing not made of parts, ‘corresponds’ to the entire activity of the brain, whatever that may be, at the moment.

Thus in James's view, consciousness is an agency that selects among conscious thoughts that are supported by unitary states of the entire brain. These original, early insights into the nature of consciousness echo with some of the more recent computational theories of consciousness, as we will see in the later parts of this chapter.

Introspective approaches like those of James came under attack in the first half of the twentieth century by the new behaviorist movement which rejected introspection since it is incompatible with objective, empirical, quantifiable methods of science. Since introspection is discarded, any mention of subjective states of mind, which can only be discovered by introspection, was also disallowed in the reckoning of science. Although there are variations of the core theme, the essential stand of behaviorists regarding the mind/brain problem was crisply expressed by JB Watson, a pioneer in the behaviorist movement, as follows:

The time seems to have come when psychology must discard all reference to consciousness; when it no longer need delude itself into thinking that it is making mental states the object of observation.

Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its methods...

Brain was treated as some sort of input/output box which produces behavior in response to environmental stimuli. Any mention of internal brain states—not just psychological or subjective states—was also deliberately avoided, which was perhaps one of the greatest weaknesses of the behaviorist approach. While it is true that the state of maturity of neurobiology of early twentieth century was simply inadequate to build meaningful neurobiologically detailed models of stimulus-response patterns of real nervous systems, in its impatience with the fallibility of introspective methods, the behaviorist approach also banished the possibility of a patient, painstaking—yet fruitful in the long-term, effort to build a theory of the brain inspired and informed by the growing knowledge of neurobiology.

The cognitive revolution which started in the ‘50s attempted to fix exactly this problem, by breaking open the behaviorist black box and studying the wheels and gears of the brain. Beginning at about the same time when the computer revolution was gathering momentum, the cognitive movement sought to express mental processes as procedures, describable sometimes in the language of mathematics, sometimes using the jargon of computer metaphor (“information processing,” “memory access,” “computation,” etc.) and sometimes, whenever it is possible, in terms of neurobiology. Cognitivist approach is identified by certain foundational principles, as Steve Pinker describes in the book *The Blank Slate*. First, the link between the mental world and the physical world can be expressed in terms of concepts like information, computation and feedback. Second, an infinite variety of behavior can be exhibited by a finite repertoire of mental programs. Third, mind is a complex system composed of many interacting parts. Another key idea of cognitive science is that the laws of the mind (like for example the laws of language and grammar) are universal and underlie minor cultural variations. The growing knowledge of the brain and nervous system also supplied the necessary substance to cognitivist thinking.

As the cognitive revolution is underway, a parallel development known as Artificial Intelligence began in the world of computers. Like the cognitive science, AI also sought to find the laws of mind and intelligence to the extent that it can be reduced to that of a machine. The basic axiom of AI is that intelligence is at its roots mechanical, expressible as a set of clearly defined rules. When the rules are sufficiently complex, the machine or the agent that obeys those rules exhibits (seemingly) intelligent behavior.

Although both behaviorism and cognitive science succeeded in recovering the bathwater of internal mental operations, they made no effort to search for the baby of consciousness still lost in the dark. The reason behind this omission was that while behaviorism refused to have anything to deal with consciousness, for cognitive science consciousness is another name given to the mental processes. On similar lines, for AI, consciousness is another name given to the computations of the brain machine. The search for the baby was called off because a shadow or a figure of

the real thing is accepted as a substitute of the real. Subjective world was denied its existence. Subjective experience is explained away as nothing more than the biophysical activities in the brain. There was a reaction against this fundamental refusal to acknowledge anything subjective or experiential from certain quarters of psychology.

Consider how Marvin Minsky one of the pioneers of modern AI dismisses the whole idea of “free will” simply as a bad idea.

Our everyday intuitive models of higher human activity are quite incomplete, and many notions in our informal explanations do not tolerate close examination. Free will or volition is one such notion; people are incapable of explaining how it differs from stochastic caprice but feel strongly that it does. I conjecture that this idea has its genesis in a strong primitive defense mechanism. Briefly, in childhood we learn to recognize various forms of aggression and compulsion and to dislike them, whether we submit or resist. Older, when told that our behavior is “controlled” by such-and-such a set of laws, we insert this fact in our model (inappropriately) along with other recognizers of compulsion...Although resistance is logically futile, the resentment persists and is rationalized by defective explanations, since the alternative is emotionally unacceptable.

Continuing in this manner of discourse, Minsky expresses the AI standpoint that “free will” is simply a matter of creating a sufficiently complex machinery, and remarks half in jest, that such machines would then struggle with insolvable philosophical dilemmas like us.

When intelligent machines are constructed, we should not be surprised to find them as confused and stubborn as men in their convictions about mind-matter, consciousness, free will, and the like. For all such questions are pointed at explaining the complicated interactions between parts of the self-model. A man’s or a machine’s strength of conviction about such things tells us nothing about the man or about the machine except what it tells us about his model himself.

As an example of a simple, straight-from-the-heart reaction to the above stance of AI, we quote American educationist John Holt. In his *How Children Learn*, an educational classic of the ‘60s, Holt responds to the above passage from Minsky as follows: “What is most terrible and terrifying about this cool, detached, witty voice... is the contempt it expresses for the deepest feelings we humans have about ourselves. His argument is a perfect example of what Laing, in *The Politics of Experience*, called the ‘invalidation of experience.’ In the words quoted above, Minsky tells us that our strongest and most vivid experiences of ourselves are not real and not true, and tell us nothing about ourselves and others except our own delusions, and that in any case he and his colleagues will soon make machines that will ‘feel about themselves’ exactly as we do. His message could be summed up: ‘You cannot learn anything about yourself from your own experience, but must believe whatever we experts tell you.’”

The AI position on “free will” in the above paragraphs is difficult to refute, not because it represents a scientifically and empirically verified truth. This summary rejection of “fee will” as an illusion is as much an arbitrary belief, or at best a hypothesis, as its contrary view that holds that “free will” exists because it appears to exist in our personal subjective experience. What is lacking is an objective, scientific

framework that considers the problem of consciousness in all seriousness that the problem deserves, and settles the issue one way or the other, as was done in case of earlier long-standing problems of science, like the problem of ether, or the problem of perpetual motion machines. The AI folks had nothing that can come close to a solid scientific case to support their view, nor did those who simply believed in “free will,” sentimentally, emotionally, or even religiously, but nothing more.

But there are inherent difficulties in making progress regarding the problem of consciousness on purely scientific lines. Modern science proclaims to deal with only objective things, while subjectivity is a primary feature of consciousness. Therefore, a science of consciousness seems to be a contradiction in terms. The challenge of building a science of consciousness is therefore to surmount this “contradiction” as were other contradictions in the history of science. It is not likely that a progress in this knotty issue is possible if we begin to study consciousness at large with its thousand modes and manifestations, just as a primary lesson on mechanics cannot begin with the n-body problem. One has to begin cautiously, take up a simple, specific, and well-defined problem, some sort of a simple pendulum of consciousness, and make progress in small steps. A lot of progress had been made in the area of visual awareness, which is certainly an aspect and operation of consciousness. Let us consider some of the highlights of research in this domain.

## **Seeing—Consciously or Otherwise**

During the course of his classic brain stimulation experiments, when Wilder Penfield stimulated visual cortices of his patients, they reported seeing simple visual patterns such as “stars lower than the bridge of [the] nose and over to the right” or “stars [which] seemed to go from the midline [of the visual field] a little across to the right.” Stimulation of other sensory cortices also produced corresponding sensory experiences of a varied kind. Thus, the fact that electrical stimulation of the visual cortex produces visual awareness, without the involvement of any explicit visual stimulus, was known for a long time.

Electrical stimulation experiments of the visual cortex have been performed for a long time as an attempt to restore sight to the blind. When minute currents of the order of microamperes are passed through the visual cortex, the subject often reports flashes of light known as “phosphenes.” In an early experiment of this kind from the ‘60s, a 52-year-old blind man was fitted with electrodes implanted near the occipital pole, a part of the primary visual cortex which receives a major part of the visual information from the retina. Stimulation at a single electrode often produced a single phosphene, a single spot of white light at a specific point in the visual field. Stimulation of some electrodes produced more than one spot, sometimes forming a whole cloud of them. Stimulation of electrodes too close to each other produced a single extended spot, but when the electrodes are about 2–4 mm apart, two distinct phosphenes were seen. When several electrodes were stimulated simultaneously, the subject saw complex predictable patterns. Usually, phosphenes disappeared as soon as the stimulation was

stopped but sometimes, after sufficiently strong stimulation, the phosphenes lingered for more than 2 min. Another interesting finding of the study revealed how different subsystems of the visual system process retinal information in different ways. During voluntary eye movements, phosphenes were seen to move with the eyes, i.e., the bright spots moved wherever the subject looked. But vestibular eye movements, a kind of reflexive, involuntary eye movements made in the opposite direction of the head movement so as to stabilize the images on the retinas, phosphenes remained fixed in space. These preliminary data from electrical stimulation experiments of the visual system clearly demonstrate that what is essential for visual experience is not so much the existence of whole wide visual world, but simply the presence of electrical activity in appropriate areas of the brain. In short, when we see, what we experience is not the visual world out there, but simply our own brain.

Similar effects were observed when magnetic pulses were directed at the visual cortex, using a technique known as Transcranial Magnetic Stimulation (TMS). Pulses of magnetic field shot at the visual cortex induce currents in the local cortical tissue, producing phosphenes as it happened in case of electrical stimulation. But magnetic stimulation in normal subjects also produced negative visual phenomena, which consist of interfering and suppressing ongoing visual experience, in addition to positive visual phenomena, which refer to the appearance of phosphenes. When strong magnetic pulses were applied, the subjects felt a transient disappearance of the retinal image. This ability of a strong TMS pulse to transiently block ongoing, local brain activity is often exploited to study the role of a given cortical area in a specific brain function.

But we do not even need these sophisticated and difficult procedures to realize that visual experience requires simply brain stimulation in the relevant areas with the visual world nowhere in the picture. All of us might have had the annoying experience of seeing a strong flash of light coming from an automobile falling in our eyes, leaving us in a daze for a few moments. The afterimage of the strong light persists, even after the stimulus itself is gone, distorting or even blocking the scene that is actually present in front of the eyes. We can very well imagine that, like in the case of a strong electrical stimulation that produced a 2 min long phosphene, the strong light stimulus must have left a longer lasting imprint in the visual cortex, which was upsetting the real-time visual experience of the observer.

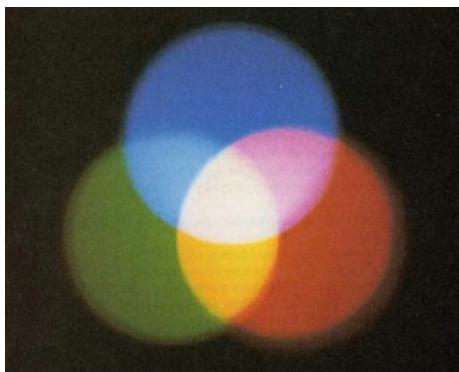
The afterimage in the above instance is merely a strong, crude phantom of visual experience. But there are other afterimages that are subtle, precise, and startlingly detailed. Just stare at the black-and-white picture shown below for about 20 s and shift your gaze towards a blank white screen (a white wall should do). You will notice the afterimage of the incomprehensible picture below turn into a very familiar face! (Fig. 12.1). Basically, the afterimage that you see is a negative of the real image shown below. The negative is incomprehensible which suddenly makes sense because the afterimage is the negative of the negative. Another interesting observation that you can make for yourself is that the afterimage moves with you as you shift your gaze around to different points of the white screen. This is analogous to the manner in which phosphenes moved along with the gaze, in the electrical stimulation experiment described.

**Fig. 12.1**

A black-and-white picture with an interesting afterimage. To see the afterimage, just stare at the black-and-white picture shown below for about 20 s and shift your gaze towards a blank white screen

**Fig. 12.2**

A map that depicts the primary colors red, green, and blue and their complementary colors, cyan, magenta, and yellow



Afterimages work with color too. Stare at the “birds and a cage” picture shown below. In the afterimage of the birds, you will notice the green bird eerily turning into a magenta bird, while the red one turns to cyan. The reason behind these phantom-like color transformations is that red and cyan are complementary colors, just as green and magenta are. Red, blue, and green are perceived as three fundamental colors by our visual system, which uses three different kinds of photoreceptors—cones to be precise—which respond best to these three colors. The three colors when mixed in equal proportions produce white (Fig. 12.2). Mixing of blue and green yields cyan, which when combined with red restores the white (Fig. 12.3). Therefore, red and cyan are complementary because each is produced when the other is subtracted from white. Similar is the case with green and magenta. It is possible to produce more complex and more startling afterimages but the take-home lesson is the same. What we visually experience is not necessarily what comes from without. If activity appears in visual cortex by whatever route, you see it.

There is another simple but intriguing phenomenon of visual perception that is often presented as a demonstration that what the mind sees is not always what is out there, but what the brain creates. There is an area called a blind spot, a small spot



**Fig. 12.3** In the afterimage of this picture, the green bird turns to magenta, and the red to cyan

in the visual fields of each of our eyes, where the eye does not receive any visual information. This deficit arises due to the absence of photoreceptors in a spot in the retina through which fibers of the optic nerve exit the eye. Since the photoreceptors form the bottommost layer among the multiple layers of the retina, fibers of the optic nerve which leave the neurons of the topmost layer, crawl along the surface of the retina, punch through a spot in the retina and exit from behind the eye. The point where this exit takes place is the blind spot, which is located at about  $6^\circ$  separation of the visual angle from the midpoint of the retina called the fovea.

To get a sense of where your blind spot is located you can try a simple experiment. Shutting one of your eyes, fixate on the cross seen below with your open eye. If you adjust the distance between the cross and your eye, you will find that at a critical distance when the “cross and spot” pattern in the picture subtend about  $6^\circ$  at your open eye, you will find, to your astonishment, the black spot disappearing from your sight. Remember that during the entire procedure you should continue to fixate on the cross with your open eye. The black spot disappears when its image falls exactly on the blind spot on the retina. When the black spot disappears, what is it that appears in its place? The answer to this question depends, curiously, on what surrounds the blackspot.

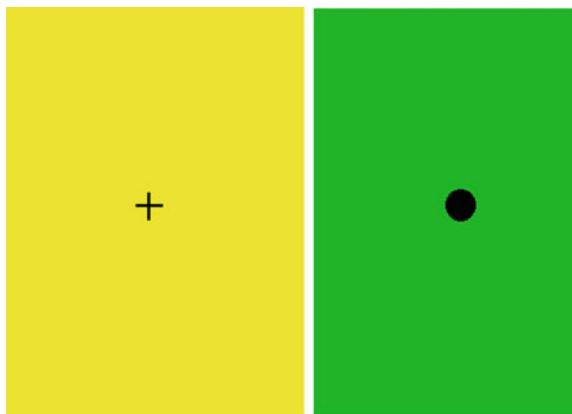
V. S. Ramchandran describes a number of interesting experiments that can be done with the blind spot which demonstrates how brain makes up new information by filling in what is missing in the blind spot. These experiments are designed in a graded fashion once the blind spot is located using the simple visual pattern of Fig. 12.4. In Fig. 12.5, the cross and the black spot are located in the middle of a yellow and a green area, respectively. When you adjust the distance of the image from your eyes such that blackspot falls on the blind spot as was done above, what you are most likely to see in place of the blackspot is not just a white hazy patch; you could see the surrounding green filling into the area that was occupied by the blackspot.

A more sophisticated version of the above experiment may be considered. Instead of a single property that can fill in, like the color green above, what would happen if there are multiple properties, like for example, a straight line passing through the blackspot? Would the green color fill in, or the straight line fill in cutting through the blackspot? When the experiment is done with the image shown in Fig. 12.5, you

**Fig. 12.4** A simple diagram for locating blind spot



**Fig. 12.5** When your blind spot is placed on the black disk on the right, the disk disappears and gets filled by the surrounding green



would very likely find the green filling in as usual, but not completely. You might also find the vertical line extending through the green which has filled into the blackspot area.

The blind spot phenomenon is not just a curious property of the visual system. On a closer examination, we would find it exposing a somewhat disconcerting property of the brain that conveniently inserts its own data wherever it is missing. Thankfully this filling in ability is rather limited in scope. Had the blind spot been larger, what would be seen is not filling in but probably just a dark patch, as though a part of the retina is truly blind.

V. S. Ramchandran and Richard Gregory describe a number of interesting filling in experiments involving moving or changing stimuli also. In one such experiment, a patient was shown a visual pattern of twinkling noise, with a large number of random dots going on and off, on which a small gray square of about  $1.5^\circ$  of visual angle was superimposed. The gray square was placed exactly  $6^\circ$  away from fixation point, so that the gray square falls on the subject's blind spot. When the patients fixated on the image for about 10 s, they noticed that the twinkling dots filled into the gray square. It would be of interest to note that the filling in did not occur instantly. On an average, calculated over several subjects, it took about 5 s for the filling in to occur, which indicates that the filling in is an active, dynamic process probably driven by the neural dynamics of visual cortex.

But then how would filling in work out when there are two properties to be filled in? To test this, the subjects were presented with random dot twinkling patterns in which black dots flashed on and off over a pink background. This time, however,

instead of the gray square which was made to fall on the blind spot, there was a pink square on which dots moved horizontally. In other words, the stimulus that fell on the blind spot was a gray square with horizontally moving dots; the surrounding region had a pink background with twinkling dots. Therefore, the dynamic texture of twinkling dots and color (pink) in the surround is in competition with the dynamic texture of (moving dots) and gray background in the blind spot area. Again filling occurred but in a more complex fashion—in two distinct stages. In the first stage, the color filled in; the pink in the surrounding area filled into the gray square. After a few seconds, the moving dots in the gray square also vanished and in their place the twinkling dots appeared.

The blind spot experiments performed by Ramachandran and Gregory were studied as an artificial version of a real visual deficit known as a scotoma. Scotomas represent local blindness caused by small lesions in visual pathway or in the visual cortex. Scotoma patients normally experience darkness in part of the visual field corresponding to the lesions. But sometimes, these patients also report remarkable filling in effects, just as in the case of the blind spot experiments described above.

Scotoma patients are practically blind in the scotoma area. Blindness may be classified into two varieties. The first one, the one that is commonly identified with blindness, is the total absence of visual capability due to damage to the eyes. There is another kind of blindness in which the eyes are intact, but the person cannot see or does not have an awareness of seeing, due to damage higher up in the visual system, say in the visual cortical areas. (Cortical damage often happens due to stroke, a condition of ruptured cerebral blood vessels.) Such blindness is called cortical blindness.

Cortically blind people sometimes exhibit an uncanny visual *ability* in the absence of visual *awareness*. For example, a patient might deny having any visual awareness when an object is placed in the patient's scotoma. But when encouraged to guess, the patient might be able to identify the object at high above chance levels. This stunning psychic-like ability is given a paradoxical name of blindsight. The name owes its origin to Lawrence Weiskrantz and colleagues who performed some of the early studies on this phenomenon in the '70s.

What the phenomenon of blindsight tells us is that it is possible for the brain to receive information from the world and even act upon it with neither our awareness nor approval. But then this should not come as a profound shock since a lot of bodily functions are conducted by the brain with neither our awareness nor approval. For example, if changes in blood glucose levels are constantly brought to our awareness and respond faithfully to our conscious approval, it might give us a lot greater control over diabetes. The shock, if there is one, is due to a shattered belief that vision cannot be an unconscious capacity like blood glucose regulation. What has always been believed to be a purely conscious faculty transformed itself under the conditions of specific brain damage into an unconscious faculty. The possibility of dual regulation (conducting an operation both consciously and unconsciously) of vision indeed comes as a surprise for anyone with a sufficiently long familiarity with how vision works. But here too, there are physiological functions that come under the "dual-regulation" category, though we may not be usually conscious of them in that fashion. A simple example is respiration. Normally, the autonomic nervous system

controls the rhythms of respiration at about 15 cycles per minute, but it is always possible to switch from the usual “cruise control” of respiration, to “manual” and consciously control the rhythm. Other rhythms, like the heart’s beat, do not easily come under conscious control. But it has been shown that by special training, it is indeed possible to exert a certain level of willed and conscious control on cardiac rhythm. What the preceding examples remind us is that conscious and unconscious performances are not watertight compartments. One can transform into the other in special conditions, in certain cases, and blindsight is one of them.

How does blindsight occur? The very fact that the brain is able to act upon visual information, consciously or otherwise, shows that information is reaching certain parts of the brain, which however, for some unknown reason, are not accessible to consciousness. The first cortical stopover of visual information is the primary visual cortex, which is the site of damage in a lot of blindsight cases. But it is known that some branches also find their way directly to higher visual cortical areas, which might be supplying the visual information necessary for blindsight performance. Another possible target is a structure called superior colliculus. This structure is located in the midbrain and receives direct inputs from retina, in addition to many other sources. Superior Colliculus is mainly involved in controlling eye movements. Another key brain structure that has been implicated is the thalamus, an important sensory hub in the brain, often described as the gateway to the cortex. Thalamus receives extensive feedback from the cortex and the thalamocortical interactions have been implicated in consciousness-related phenomena. But there is no final theory of blindsight as on date, just as there is no final theory of consciousness yet.

The visual phenomena discussed so far in this chapter—afterimages, blind spots, blindsight, and all—demonstrate a nonunique and variable relationship between what is out there, or what enters the eye, and what is experienced higher up in the brain/mind or the mind-related parts of the brain. But the picture is not complete, and not sufficiently precise, unless we are able to actually measure brain’s activity, in the form of neural spike activity, and establish their connection to conscious experience. If the visual experience does not strictly correlate with the externally presented image, where does it come from? What exactly determines visual awareness? It is very reasonable to expect that something that goes on in the brain precisely determines what is experienced, though brain’s activity ought to also reflect the external stimuli. If you applied the above logic to the problem of scotomas or the blind spot, there must be someplace in the visual cortex which ought to have received inputs from the blind spot area, but actually does not; therefore neural activity in that part of the cortex would indicate that the relevant information from the retina is missing. But somehow, somewhere else perhaps in the brain, filling occurs which is experienced by the consciousness; there must be neurons whose activity reflects the filling in. This scenario, should it turn out to be true, throws up an interesting question: what determines what kind of neural activity enters consciousness and what kind does not?

A line of research in the area of visual awareness had precisely addressed the above question. This line of research dealt with an interesting phenomenon called binocular rivalry. We all know that when we look upon the world with our two

**Fig. 12.6** A stereoscope

eyes, we grab two slightly different images of the world, taken from two slightly different vantage points. Our brain combines these pairs of images and constructs a three-dimensional image of the world. The two images that fall on our two retinæ are actually not very different. In a sense, one may be constructed by slightly and locally shifting patches of the second image. This is the reason why it is possible to “fuse” the two images and unearth the depth information buried in them. The redundancy that goes with the common features in the two images enables us to compare, work out the correspondences between the two images, and compute depth. But the trouble, or the fun, depending on how you react to it, begins when the two images are artificially constrained to be dissimilar. When the two images are made to be dissimilar, producing what is known as binocular rivalry, the brain, unable decide what the impossible object is to which this strange image pair corresponds to, throws illusions at the observing consciousness, showing it scenes that do not strictly match with the static images shown to the brain.

The phenomenon of binocular rivalry was supposed to have been known for centuries with the earliest reports dating back to 1593. The enterprising inventor of this quaint pastime, a person named Porta, was credited with an unusual ability. He used to be able to put one book in front of one eye, and a second book in front of the other, and read the two books alternately shifting his attention (he used the phrase “visual virtue”) from one book to the other. A large number of interesting studies have been performed for several centuries since the effect was first discovered. But there were practical problems involved in performing these experiments. Looking at two objects with the two eyes, making sure that each eye sees only one of the objects and not the other, is not easy. But the practical problem was solved in 1838 when Charles Wheatstone invented the stereoscope, a convenient device in which two different images can be projected to the two eyes separately (Fig. 12.6).

The key and unexpected observation that is noted in binocular rivalry is that when two different images are presented, consciousness does not perceive an incongruous

mixture of two incompatible images. It actually sees the two images one after another, with a rhythmic alternation. Some studies have also closely observed the manner of transitions from one image to the other. A trigger for change begins at one point of an image, and from thereon it spreads, like a forest fire, over the entire visual field, changing one pattern into another. Like an ocean wave rolling over the shore, erasing a pattern written on the sands, one pattern is seen to gradually spread over another erasing it.

What happens in the brain as this battle between percepts goes on in the consciousness of the viewer? In order to answer this question, Sang-Hun Lee, Randolph Blake, and David J. Heeger took functional Magnetic Resonance Images (fMRI) of subjects who were presented with binocular patterns. The patterns had circular symmetry and consisted of black-and-white stripe patterns. One of the patterns had radial black-and-white patterns present only within an annular portion of a disc. The second pattern had identical organization, with the only difference that the contrast of the second pattern is much weaker than that of the other. When the patterns were presented separately to the two eyes, the subjects reported seeing, as expected, wave-like transitions from one percept to the other. fMRI recordings of the primary visual cortex (V1) taken during this process detected similar sweeping waves traveling over primary visual cortex, V1, as the subjects observed binocular patterns. The subjects were also asked to indicate when they observe a pattern transition. Using this feedback, it was possible to measure the frequency of transitions. This frequency is found to match with the frequency of traveling waves over V1. These results strongly indicate that the visual experience, though at apparent variance from the static visual stimulus presented, is caused by actual neural activity in V1.

Why does the brain respond with a dynamic pattern to a static visual stimulus? We know that both eyes receive visual information from both left and right visual fields. However, the fibers of the optic nerves from the two eyes are routed to the brain such that information about the left visual field in both the eyes is collected and brought to the right primary visual cortex. Similarly, information about right visual field is brought to the left primary visual cortex. However, neurons in the primary visual cortex tend to respond preferentially to input coming from only one eye—left or right—a property known as ocular dominance. Neurons of a given side (right or left) of V1 that respond to right and left eyes are not spatially segregated but form intricate, interpenetrating regions, resembling zebra stripe patterns, known as ocular dominance maps.

It is also known that V1 has several varieties of neurons, some excitatory and some inhibitory. The spiny (referring to the appearance of the cell body) neurons of V1 have local excitatory connections, while the smooth neurons have local inhibition. Taking advantage of the above physiological data regarding V1, Sidney Lekhy proposed that mutual inhibition between neurons that respond to inputs from distinct eyes can produce oscillations. This idea has been elaborated by several other researchers who proposed models that can explain several aspects of binocular rivalry including frequency of transition, velocity of traveling waves, effect of noise on percept transition, and so on.

Studies on binocular rivalry have also been performed on monkeys, with the added advantage that it is generally easier to take single neuron recordings from a monkey's brain than in case of human subjects. Nikos Logothetis and Jeffrey Schall presented moving gratings to a monkey, so that one eye was shown upward moving bars, while the other eye was downward moving bars, producing rivalry. In case of human subjects, the experimenter comes to know about the current percept by the feedback from the subject. Likewise, the monkeys were trained to signal whether they were seeing upward or downward moving grating. Based on this feedback, it became clear that the monkey also saw alternating percepts just like humans.

Since the stimuli in the above case consisted of moving patterns, neurons in a visual cortical area known as Middle Temporal area (MT) responded significantly. (MT neurons are known to process moving visual patterns). But the key question here is: do these neurons respond to the external stimulus or the percept? Single neuron recordings showed that both are true. Some neurons responded to the stimulus actually presented, while other correlated with the percept that the monkey is actually experiencing. This data indicates or reconfirms a very important truth regarding neural correlates of visual awareness. All neural activity, even that of visual cortical areas, does not necessarily contribute to visual awareness. Which neural activity contributes to visual awareness, and which does not continue to be an unanswered question.

Similar findings have been echoed by studies on human subjects which considered activity in higher visual cortical areas beyond V1. Lee, Blake, and Heeger also directed their fMRI-based studies of binocular rivalry to higher visual cortical areas, like V2 and V3, and found traveling waves there too. Waves in these higher areas were found to be dependent on visual attention. There are two possible ways of visually attending to a stimulus: by actually turning the eyes towards the target, which is known as overt attention, or by inwardly directing attention to the target even though the eyes fixate on a different location, which is known as covert attention. In the above fMRI study, the subjects were asked initially to pay (covert) attention to the point where one pattern changes into another. Later when the subjects were asked to divert the attention away from the changing percepts, wave activity in V2 was attenuated, while the wave pattern of V3 actually reversed. But the waves in V1 persisted even when attention is diverted. Thus, traveling waves in V1 did not seem to be affected by attention. But the most interesting aspect of the above study is that the presence of waves in V1 is necessary but not sufficient for the subject to see binocular rivalry; wave activity in higher visual areas must accompany similar activity in V1. This brings up the question of the possible origins of traveling waves of V1, and of the more stringent correlates of binocular rivalry. It is possible that the V1 waves are strongly influenced by higher areas, like the prefrontal for example, which modulate the V1 waves by feedback influence. Though the precise details of this top-down influence are not fully known, what is clear is that V1 is not the sole site of binocular rivalry (and perhaps of visual awareness in general) which is controlled by a hierarchy of not only visual areas but also from much higher control centers in the prefrontal. Working out the full details of this control is likely to occupy research on visual awareness for a long time to come.

## On Being Aware of Being Touched

Although it seems sometimes that vision hogs too much attention (perhaps greater in proportion than that of the visual cortex in the entire sensorium of the brain), it is not that consciousness is an exclusive prerogative of the visual sense. Naturally, every sensory power has its own unique consciousness or sensory awareness. Every sense reports ultimately to its unique, conscious *Senser* and *Knower*. Tactile sense may not have the glamor enjoyed by vision but it is definitely important, and in some respects more important than vision for a very simple reason. It is possible to shut off other senses: vision by shutting eyes, audition by plugging ears, smell by shutting nostrils and taste by shutting the mouth. But it is not easy to free oneself from all tactile stimulation even for a moment. Even when you are not touching anything in particular with your hands, the very basic act of standing or sitting or lying down involves contact with another surface and therefore tactile stimulation. Even in a much more exotic state of freely floating in space, you cannot help feeling your clothes (most probably heavy and unwieldy). And even when you are free floating in space without clothes, for whatever unearthly reason, you still cannot help feeling your body, since that is the primal, fundamental tactile sense—called *proprioception*—that you cannot get rid of unless you are unconscious. Perhaps because it is all the time present, we tend to ignore it. Perhaps it takes an unfortunate deprivation of other “prominent” senses to celebrate touch and proclaim, like Helen Keller, that “a lush carpet of pine needles or spongy grass is more welcome than the most luxurious Persian rug.”

The journey of tactile information to the brain begins in the skin, where there is a rich variety of touch transducers that convert tactile information into electrical potentials. Some of them that are located superficially detect light touch, and those located deeper in the skin pick up deep touch and pressure. There are some that respond best to vibrating touch. The hair on the skin also contributes to tactile experience, as when we experience the joy of a cool zephyr passing over our skin. Humans may not think much of hair as a source of tactile information, but whiskered animals, like rodents, for example, make profitable use of the news that their hirsute (a serious medical term for “hairy”) surfaces convey to their brains. All the above sources of tactile information are bundled under the general category of cutaneous information.

Another category of information, which is tactile-like, since it is based on mechanical sources, comes from the muscles and joints. There are sensors in the muscle that convey data about muscle length, rate of change of that length and muscle tension to the brain. There are also special sensors in the joints that convey data about joint angles. All these types of information inform the brain about the status of the muscles and joints, which constitutes the so-called “proprioception” or, in plain English, the “position sense.”

It can be debated whether touch is more important than vision, or less, but it is certain that touch is similar to vision in many ways. First, the transducing system in both cases consists of a two-dimensional sheet of sensors—the retina in vision, and the skin for touch. This similarity has been brilliantly exploited by Paul Bach-y-rita in

his sensory substitution experiments, the prominent among which consists of using touch to restore vision to the blind. In one of these devices, the blind subjects were fitted with a  $20 \times 20$  array of stimulators attached to their back. The stimulators, called tactors, could be electrical, which pass minute currents into the skin, or mechanical, which vibrate and produce a tingling sensation. A low-resolution camera is fitted to the head of the subjects, facing the direction in which the subject is facing. Image from the camera is converted into electrical signals and sent to the array of tactors fitted to the back of the subjects. Thus, the optical image from the camera is converted into a tactile image which is felt by the subjects as a pattern of tingles on their back. The subjects were able to discriminate simple patterns of stimulation after a few hours of training. But in the early stages, they would stay focused on their backs in order to decipher the strange “visuo-tactile” information that is flowing into their nervous systems. But with practice, they stopped paying attention to their backs and shifted attention to what is in front of them, as if they can “see” what is ahead of them. It is as though their awareness, which was tactile until that stage, has transformed into a new vision-like awareness, which is “looking” at what is in front of them. When the sensors were fitted on their abdomen, the subjects were able to quickly readapt and work with the new arrangement. Even with the low-resolution images that the subjects had to work with, they were able to effectively navigate their personal space.

How is this tactile “seeing” different from the regular seeing done with retinas and visual cortices? The key difference is the huge disparity in the resolution—1 million fibers in each optic nerve as opposed to a  $20 \times 20$  array of sensors. Some young subjects who were fitted with Bach-y-rita’s device complained that they did not feel the affective arousal that would be natural to someone of their age, when they were shown pictures of beautiful women. Philosopher Daniel Dennet compares this tactile vision with blindsight in his book *Consciousness Explained*. Blindsight subjects usually report that they cannot see anything. When urged to describe whatever little that they might be “seeing” these subjects too report seeing *something*. Dennet quotes from the original blindsight studies by Weiskrantz (1988) on this matter [pg 338]:

DB [a blindsight subject] sees in response to a vigorously moving stimulus, but he does not see it as a coherent moving object, but instead reports complex patterns of ‘waves.’ Other subjects report ‘dark shadows’ emerging as brightness and contrast are increased to high levels.

Dennet remarks that this seeing of blindsight subjects is perhaps similar to that of Bach-y-rita’s patients. In both cases, it is a reduced form of vision, a milder form of visual or vision-like awareness. Since the awareness is so weak, it is difficult to definitely label it as visual or tactile. The experiment indicates that it is perhaps a blunder to consider different sensory modalities as if they occur in watertight compartments. Perhaps there is just one awareness whose intensities and nuances are experienced as different forms of sensory experience.

Bach-y-rita’s devices show how tactile sensation can be converted into visual sensation. But there are some simple ways in which tactile sensation can be radically altered, or even extended beyond one’s body, as was demonstrated in an experiment by Matthew Botvinick and Jonathan Cohen. In this experiment, the subjects were

asked to rest their left arm on a table. A screen was placed between the subjects' eyes and the arm so that the left arm remains hidden from their view. A rubber arm was placed visibly on the table in front of the subjects. Now the experimenters stroked rhythmically, using two small paint brushes, the subjects' real left arm on one hand, and the rubber hand on the other. It was ensured that the rhythms were in perfect synchrony. The subjects were then interviewed about what they felt. The subjects' reports were startling: "It seemed as if I were feeling the touch of the paintbrush in the location where I saw the rubber hand touched"; "It seemed as though the touch I felt was caused by the paintbrush touching the rubber hand"; "I felt as if the rubber hand were my hand." Some even felt momentarily that they had two left hands.

The strong synchrony between the rhythms of stroking given to the real hand and the rubber hand plays a crucial role in creating this illusion. The illusion was rendered significantly weaker when a small asynchrony was introduced in the two rhythms. In the current situation, the brain is receiving two streams of information—the tactile from the real hand and visual from the rubber hand. The two streams are in synchrony, a property that seems to create the illusion that the two sources of synchronous data streams actually are one and the same, or closely related, or belong to the same person. Synchrony seems to promote a sense of identity.

This observation is strongly analogous to a finding from vision research by Wolf Singer and Charles Gray. A question that motivated the research of Singer and Gray is as follows. When an extended visual object is presented to the brain, different neurons in distant parts of the brain respond to the object. For example, if it is a bar of a finite length, and a specific orientation, visual cortical neurons that are tuned to that orientation respond to the bar. But how does the brain know that all these neurons are responding to the same object, the same bar? One possible solution to this question is that there are neurons in higher visual cortical areas that respond to combinations of properties detected by neurons of lower layers. In other words, if the neurons of lower layers respond to parts of the bar, neurons of higher layers respond to the entire bar. But such a solution requires an impractical explosion of possible objects, each representing a combination of properties. There must be more an economical way of binding the various features of a single object and producing in the brain the sensation of experiencing a single object. In the experiments of Gray and Singer, visual stimuli were presented to kittens with multiple electrodes implanted in their visual cortex. Firing patterns in different parts of the visual cortex were recorded by the electrode when single visual objects were presented. For example, when a moving bar pattern of a given orientation was presented, it was observed that neurons that responded to the bar fired with significant levels of synchrony. Particularly, the synchrony occurred in the range of 40 Hz, an important frequency that falls in the so-called gamma range of brain rhythms. Thus in this case, synchronous activity of neural firing patterns coded for the idea that the activities of those neurons represent the same object. The idea that neurons that respond to a single object could be firing in synchrony, thereby obviating the need for the existence of separate object specific neurons that combine features of the object, was first suggested by Christoph van der Malsburg. The observations of Singer and Gray bore evidence to this elegant suggestion. Use

of temporal synchrony to bind different features of the same object, and represent the unitariness of the object, is known as “temporal binding.”

There is no data available regarding the possibility of temporal binding, most likely in the somatosensory cortex, of subjects experiencing rubber hand illusion. But a similar experiment by Claudio Bablioni and colleagues demonstrated a strong synchrony in the frequency band of 36–44 Hz in the somatosensory cortex in response to stimulation of contralateral thumb. For this purpose, they used a technique that measures the tiny magnetic fields generated by currents produced by neural activity. It will be interesting to extend such studies to human subjects experiencing the rubber hand illusion.

A large number of experiments in both visual and somatosensory domain demonstrate that synchronized gamma oscillations represent response of neurons to a single object. Extending the relevance of gamma oscillations from integrity of object representation, to integrity of object perception, Christoph Koch at Caltech and Francis Crick, co-discoverer of the double helical structure of DNA, have suggested that these oscillations might be neural correlates of conscious perception. That is, synchronized gamma oscillations not only code for the integral representation of an object; they are also an essential requirement for the subject to have a conscious perception of the object.

There is some evidence that supports the “astonishing hypothesis” of gamma oscillations being the substrate for conscious perception. A study conducted by Lucia Melloni and colleagues on conscious perception of visually presented words is a good case in example. Stimuli presented both consciously and unconsciously or subliminally can activate nearly the same areas of the brain, including the higher cortical areas. Stimuli that are presented too briefly, or are masked quickly by a distractor, are not perceived consciously, but there is evidence that brain responds to these stimuli. What then distinguishes stimuli that are consciously perceived from those that are not? The Melloni study demonstrated that when a stimulus is perceived consciously, there was a long-range but transient synchronization in the Electroencephalogram (EEG), in the gamma frequency range. Deficits in synchronization in high-frequency bands have been found in neural and psychiatric disorders like Schizophrenia, Parkinson’s disease, Autism, and Epilepsy.

## The Subjective Timing Experiments of Benjamin Libet

Just as there are conditions on neural activity for it to enter consciousness, there are also conditions on stimulation for the subject to become conscious of it. Even from the earliest days of Penfield’s stimulation experiments, it was known that direct electrical stimulation of sensory cortices produces the corresponding sensation. More detailed studies of cortical stimulation of the sensorium by Benjamin Libet yielded some thought-provoking and controversial results, with deep consequences for our understanding of the conditions of conscious awareness. Libet considered the conditions of stimulation parameters for the subject to feel the conscious sensation. First,

it was found that the stimulation current has to exceed a threshold value to produce a conscious sensation. Second, the current has to be delivered for a minimum duration to produce conscious sensation. Third, when the stimulation is in the form of pulses, it was observed that when pulse frequency is higher, it required smaller current levels to produce the threshold sensation. As these three parameters—current intensity, stimulation duration, and pulse frequency—are varied and the complementarities among the three parameters involved in reaching the threshold of sensation were assessed, Libet noticed an invariant quantity. Recall that if the stimulation duration is increased, the sensation can be produced with smaller current levels (assuming the pulse frequency is constant). However, the current level cannot be reduced indefinitely; there is a minimum current required, at a given pulse frequency, to produce conscious sensation. At that minimum intensity—what Libet calls the liminal intensity or Liminal I—the stimulation duration turns out to be about 0.5 s. This minimum duration of 0.5 at Liminal I was found to be necessary for a range of pulse frequencies. That is, at the threshold current levels, the subject is not aware of the stimulation unless it is applied to the cortex for about 500 ms.

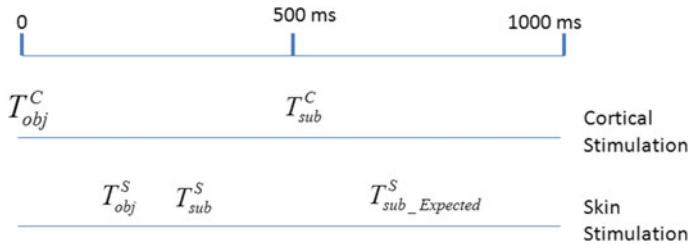
A similar result obtains when the stimulus is given peripherally, to the skin. In this case, even when intensity of stimulus is sufficiently strong, it takes about 500 ms of stimulation for the subject to have a conscious sensation of it. The curious aspect of this result is that it takes only about 20 ms for the peripheral signal to reach the cortex. Why did it take so long for the subject to feel the stimulus, or to borrow the vivid language of Daniel Dennett, “why did it take so long to reach *you*,” though it reached your brain long ago? Just as in the case of cortical stimulation it appears that the effect of the stimulus must reach a state of adequacy, for the subject to detect it, and to reach adequacy takes almost half a second.

Therefore, in both cases (cortical stimulation and skin stimulation) it takes about 500 ms for the subject to feel it. Now let us define a few quantities for ease of reference.

|                                       |  |
|---------------------------------------|--|
| $T_{\text{obj}}^{\text{C}}$           | is the time at which the cortex stimulus is given in the objective world               |
| $T_{\text{sub}}^{\text{C}}$           | is the time at which the cortex stimulus is felt in the subjective world.              |
| $T_{\text{obj}}^{\text{S}}$           | is the time at which the skin stimulus is given in the objective world.                |
| $T_{\text{sub}}^{\text{S}}$           | is the time at which the skin stimulus is felt in the subjective world.                |
| $T_{\text{sub\_expected}}^{\text{S}}$ | is the time at which the skin stimulus is expected to be felt in the subjective world. |

From the preceding discussion, we would assume that  $T_{\text{sub}}^{\text{C}} - T_{\text{obj}}^{\text{C}}$  is about 500 ms, and  $T_{\text{sub}}^{\text{S}} - T_{\text{obj}}^{\text{S}}$  is about 500 ms. The first statement ( $T_{\text{sub}}^{\text{C}} - T_{\text{obj}}^{\text{C}}$  is about 500 ms) is true but the second ( $T_{\text{sub}}^{\text{S}} - T_{\text{obj}}^{\text{S}}$  is about 500 ms) turns out to be untrue. It was found that there is hardly any delay between  $T_{\text{sub}}^{\text{S}}$  and  $T_{\text{obj}}^{\text{S}}$ , whereas it would be expected that the delay in this case too would be about 500 ms.

We have been glibly talking about timing of subjective events, but unless the quantity can be defined clearly and measured objectively, it has no meaning. How can we objectively measure something essentially subjective? How did the experimenter find out when the subject felt the sensation? Obviously, it is not the time when the



**Fig. 12.7** Subjective timing relationships in Libet’s experiment. The delay between cortical stimulation and the corresponding subjective experience is about 500 ms. Surprisingly, in case of skin stimulation, the delay is much smaller

signal has reached the brain. In order to time this subjective event, Libet’s group employed an ingenious technique. The subject is shown a circular dial on which a dot keeps rotating at a sufficiently low speed that the subject watching it can keep track of its position. When the subjective event occurs, when the subject feels that the stimulus has arrived, the subject simultaneously remembers the position of the dot on the dial at the exact same time and reports it to the experimenter at the end of the trial. Thus, the position of the dot on the dial represents the time of occurrence of the subjective event.

Now something curious happens when stimuli are given both to the cortex (C) and to the skin (S) and the subject is asked to compare and report which signal first entered his/her conscious awareness. The relative positions of the timing of the four events,  $T_{sub}^S$ ,  $T_{obj}^S$ ,  $T_{sub}^C$ ,  $T_{obj}^C$  is depicted in Fig. 12.7. The delay between  $T_{sub}^S$  and  $T_{obj}^S$  is only a few 10 s of milliseconds, while  $T_{sub}^C - T_{obj}^C$  is about 500 ms. Even when the skin stimulus was delivered a few hundred milliseconds after the cortical stimulus, the subjects felt that the skin stimulus arrived sooner. A stimulation that was given right in the brain, in the cortex, took longer to enter consciousness, than one that was given far off on the skin!

Libet explains this strange subjective timing results in terms of what he calls “backwards referral of subjective time.” The consciousness automatically refers to the subjective event backwards in time when the stimulus is delivered peripherally. Perhaps since it always takes that long for a peripheral sensory stimulus to enter consciousness, it might have learnt to refer it backwards by a fixed compensatory duration so that the delay can be ignored. The idea of subjective backwards referral of time seems quite strange, but Libet defends it as follows:

Subjective referral backwards in time is a strange concept and perhaps is not readily palatable on initial exposure to it. But there is a major precedent for it in the long recognized and accepted concept of subjective referral in the spatial dimension. For example, the visual image experienced in response to a visual stimulus has subjective spatial configuration and location that is greatly different from the spatial configuration and location of the neuronal activities that give rise to the image.

We might be watching what we believe to be a flat portrait on the wall in front of us. But that flatness is conveyed to our subjective self by the activities of a neuronal

layer, which is a part of the convolutions of the occipital lobe. It is not clear how space of this curved brain surface is mapped onto—or referred to—the space of our subjective visual world. But we do not find it mysterious perhaps because we have not thought about it or are simply used to it. Similarly, Libet argues that backwards referral of subjective timing is there for a purpose and not any more mysterious or unacceptable than subjective referral of space.

The summary of Libet's experiments can be reduced to two points:

- (1) Cortical activity, triggered by a sensory stimulus, must persist for a minimum duration in order to produce a conscious sensation. This prevalence of cortical activity for a minimum duration is known as neuronal adequacy.
- (2) Once the neuronal adequacy is achieved, the subjective timing of the experience is referred backwards in time.

Libet's experiments have triggered an immense amount of thinking and discussion not only in somatosensory research community but more fundamentally among those interested in the mind–body problem. It came as a strong challenge to the purely materialist position that consciousness is just another name to the neural activity, and does not deserve an exclusive study. But showing repeatedly, in a large number of experiments, that consciousness lives in its own time, which is certainly related to the physical objective time, but is different from it, Libet gave the problem of consciousness a validity and significance that it perhaps did not have before.

## Distortions in Doership

Thus we have seen that, in the sensory domain, first and foremost, what we are conscious is merely activity in the brain, mostly in the corresponding sensory cortex; we neither really know nor care about what is “really” out there. Even the cortical neural activity enters our consciousness only on certain conditions—higher cortical areas must also be active in addition to the lower ones, synchronized firing in gamma range and so on. There is a delay between the stimulation and its sensation which consciousness has learnt to cleverly deny. Furthermore, in sensory perception, consciousness acts as an agent of selection, just as William James had predicted, of a particular percept among several possible ones at a given moment and a given context.

The possible role of consciousness on the motor side is likely to be very different. The motor side is all about action, generating movements. In this domain, we expect that the consciousness present itself in its active form, will, which initiates movements. Will, free by its very nature, we would like to believe, is the prime mover; brain takes that primal command, improvise and embellish it according to the context, work out the work breakdown, call the necessary movement subroutines, orchestrate an extensive musculoskeletal machinery, and throw into the real world an efflorescence of movement.

But the will cannot do all this without any help. It needs to rely on sensory data that pours in from the posterior cortex and adjust its motor plans accordingly. Thus, we would expect, will inserts itself between the sensory input and the motor output, acting like some sort of a switch that closes the sensory-motor loop, or keep it open. If there is no such conscious blockage or vetoing of sensory input, and if all sensory input is allowed to progress onwards to some irresistible motor consummation, then we would be at the mercy of the world, and its sensory shocks. Without the prerogative to select from an array of possible responses to the sensory inputs from the environment, we would be forlorn playthings of the sensory world and not conscious, willing creatures that we believe ourselves to be.

Curiously, French neurologist Francois L'Hermitte found exactly such patients—people who are helpless slaves of the sensory world. These patients had partially missing frontal lobes and had an eerie manner of responding to the external world. If a feature in their immediate neighborhood prompts a certain act, they would unhesitatingly enact it irrespective of the relevance or the necessity of the act. If they saw a comb they would pick it up and comb themselves. If it is a toothbrush, they cannot resist brushing themselves. In many of these patients, this tendency took an unfortunate turn: it provoked in them criminal tendencies. If they see an unguarded wallet, they would try to steal it. If they see an unattended automobile, they would try a break in. L'Hermitte called this tendency “environment dependency syndrome.”

L'Hermitte performed a series of revealing experiments to demonstrate how this environmental dependence is expressed as extreme behavior. In one experiment, L'Hermitte invited two of his patients home. He took one of the patients into his bedroom without any prior instruction. The moment the patient saw the bed which was neatly prepared, he undressed himself, and got into the bed, ready to sleep, though it is only midday.

L'Hermitte persuaded the first patient to get up and leave the bedroom. The second patient, a woman, was now brought into the same room. The moment this person saw the crumpled sheets on the bed, she got down to the task of preparing the bed, all without a word of instruction from the doctor.

In another experiment, L'Hermitte took one of his patients to another room, and just before entering the room, he uttered the single word “museum.” From the time the patient entered the room, he began to behave as if it really was a museum and started looking intently at the paintings on the wall, one after another. At one point in the wall, there was a gap with a missing painting. The doctor deliberately placed a hammer and some nails near that gap. When the patient came to this spot, he thought that the gap had to be filled. He immediately took the hammer and started driving one of the nails into the wall.

Such extreme dependence on environmental stimuli occurs because the highest authority of decision-making that selects and permits, or vetoes, actions suggested by the external stimuli is absent. Prefrontal Cortex (PFC), the large anterior portion in the frontal lobe, is that highest authority and decides and plans the actions to be performed. One of the prime functions of PFC may be described as goal-oriented behavior, which can be considered as contrary behavior to environmental dependence seen in L'Hermitte's PFC damaged patients. In goal-oriented behavior, the subject

has a goal in his/her mind and initiates and executes a series of actions that take the subject to the desired goal. In the process, the subject may choose to use sensory feedback from the environment, but he/she is not a slave to such feedback on a moment to moment basis.

Let us consider an elementary example of goal-oriented behavior: getting a drink of water. You hold the picture of a bottle of chilled water in your mind, even though such an object is right now not in your vicinity. You navigate your way through several rooms, without getting distracted by what you see on the way, the TV, the laptop, your pet dog, or whatever—and find yourself in the kitchen. Assuming your PFC has not forgotten the goal or the purpose of why it drove you to the kitchen, you would open the refrigerator and grab your goal. Or if your PFC had dropped the ball, then you would fumble a bit, wonder why you have come to the kitchen, retrieve after a moment's struggle your goal information, and take the necessary action. If the above role of PFC is projected to more complex decision-making situations of life, it gives an idea of its key position in brain function.

PFC decides and plans but when it comes to actual execution of commands, it has to rely on motor cortex and other subcortical loops that support motor action. The highest motor cortical area known as the Supplementary Motor Area (SMA) is located above the primary motor cortex (M1), bordering on the “leg” area of M1, and extending into the medial surface where the two hemispheres come into contact with each other. SMA is known to be particularly active when a person is engaged in conscious, willed action (like lifting a hand to wave at someone), as opposed to action that is predominantly derived from and driven by sensory input (like moving a hand to catch a ball). SMA is also the prime mover in voluntary action. When you are sitting quietly, motionless and suddenly decide to lift your hand at a self-chosen moment, without any involvement of an external trigger, the SMA first gets activated, subsequent to which activation spreads to the contralateral primary motor cortex, activating appropriate movement. SMA activation is also found during conscious mental rehearsal that is not accompanied by explicit movement. In this regard, we may describe the role of SMA as one that supports the movement of thought that precedes the physical movements of the body. Another important aspect of SMA functions, one that is relevant to its role in willed action is that, when SMA is electrically stimulated, subjects report that they had the feeling of “wanting to do” something. SMA activation creates an intention of performing an act.

So far in this section, we have used the intuitive, commonsensical description of how we think willed action works. First, the will to perform a certain action arises in the brain, (probably in the form of neural activity of a higher brain area like the SMA), which then progresses to M1 and so on. But this simple universal belief in the doership of the conscious self in voluntary actions is shattered by a radical experiment by neurosurgeon W. Grey Walter in the ‘60s.

Grey Walter had his patients implanted with electrodes in the motor cortex (most probably SMA). The patients were shown slides from a carousel projector. It is possible for the patient to advance the carousel and change the slide simply by pressing a button. But actually, unknown to the patients, the button had no connection with, and therefore no control over, the slide projector which was in reality triggered

directly by the signals picked up by the electrodes implanted in the motor cortex. The objective of the experiment is to see if the activity in the motor cortex is accompanied by a sense of initiating movement, or willed action. The experiment began, and the subjects began to push the button and watch the slides move. Even after the first few button pushes, the subjects were shocked by a strange experience. The slide projector seemed to anticipate their intentions and move on its own without them actually pushing the button! Sometimes they had to withhold themselves from pushing the button just to ensure that the carousel does not advance by two slides. What is happening here?

If the sense of will is born first, followed by actual movement, the subjects would first become conscious of their intent and would observe the slide moving, even though it did so without their button push. But if the activity in the motor cortex starts first, with conscious intent arising after a delay, the motor cortical signal almost instantaneously reaches the projector and moves it; the subject would become conscious of the will associated with the cortical activity after a delay. It took some time for the motor cortical activity to enter the conscious self, just as it took some for cortical stimulation to reach “neuronal adequacy” and produce a somatosensory experience in Libet’s experiments of the previous section. The subjects therefore become aware of their own intent a bit late, by which time the movement has already begun, which they find quite startling.

The above account of temporal ordering of conscious will and movement needs to be made more rigorous by precise timing measurements of the three events present in the picture: Willed action (W), motor cortical activity (C), and actual movement (M). Our naïve ordering of the three events would be as follows:

$$W \rightarrow C \rightarrow M$$

But delicate timing experiments performed by Benjamin Libet in this regard shatter this naïve expectation and forces us to rethink the whole question of “free will” and its philosophical implications. A few words on the work on the electrophysiology of voluntary movement that paved the way for Libet’s experiments in this area. In the ‘60s, Kornhuber and Deecke took Electroencephalogram (EEG) recordings from subjects involved in voluntary actions. When the subjects were engaged in self-initiated wrist movement, a slow buildup of activity was observed in the midline central (Cz) electrode, close to the SMA, nearly a full second before the movement was initiated. The discoverers of this potential gave it a German name, Bereitschafts Potential (BP), which was substituted by Readiness Potential (RP) in the English speaking world. The BP represented brain’s efforts at preparing to move. This slow buildup was not observed when the subjects merely moved in response to an external stimulus. Careful analysis of the source of this readiness signal (which was necessary since it was a scalp recording, and not a direct one from the cortex) showed that SMA is the key contributor.

Of the three events (W, C, and M) listed above, the beginning of BP represents C and the beginning of movement, which was measured by an Electromyogram (EMG) in Libet’s experiments, represents M. (EMG is a technique that measures electrical

activity of muscles in a state of activation.) It now remains to determine W, the time when conscious will arises in the subject's mind. To measure W, Libet again used the same setup used to measure conscious timing of somatic sensation. He had a dot move on a dial at about once in 2.56 s. Subjects remembered the dot position when they first thought of their intent to move. When W was thus measured, the results obtained were quite startling. W did not precede C, as expected. The actual order of the three events was found to be

$$C \rightarrow W \rightarrow M.$$

The cortical activity began first, and the sensation of will followed after about 350–500 ms. It is interesting that in the case of willed action too, brain's activity seems to reach some sort of “neuronal adequacy” for the subject to become conscious of the intent. But that causes serious philosophical difficulties in case of willed action, which did not arise in case of conscious sensation. If conscious sensory experience followed after sensory stimulus, only with a delay that is slightly longer than expected, the result is not unpalatable. It qualitatively matches with our intuition; the mismatch is only quantitative. But in case of willed action, the finding that the sense of will occurs *after* the brain's preparations for movement are long underway, is disturbing, and tears down our fundamental cherished beliefs in our self-hood, in our privileged position as the prime mover of our willed actions. It shatters our intuitive, commonsensical expectation of how willed action works. Our brain initiates “our” actions, Libet's experiments seem to suggest—all by itself and is courteous enough to inform us “by the way” about the proceedings that were initiated right under the nose of the conscious self.

## Varieties of Consciousness<sup>1</sup>

The story of consciousness research in its modern form began with a stark denial of the very existence of consciousness by the behaviorists, and was derailed to some extent by the cognitivist attempts to explain it away in terms of “processing” in the brain. A more pragmatic approach, though slow it might seem in its pace, is more secure, rigorous and sustains hope of an ultimately satisfying theory of consciousness, seems to have been emerging over the past few decades. This approach accepts a priori that there is such a thing called consciousness and subjective awareness which can only be detected as on date by the subject's report. In the notation of Nobel laureate Gerald Edelman, we are dealing with two domains, labeled C and C'. C refers to the world of conscious experience and its contents (“I see the color blue,” “My left hand was touched before the right hand”). This world is completely private to the subject and can only be revealed to the experimenter by the subject's report. C' refers to the objective world of stimuli (visual, auditory, etc.) and neural activity

---

<sup>1</sup>To borrow from William James.

which can be measured by EEG, imaging techniques, single-cell recordings and so on. The problem of consciousness now becomes the problem of working out the myriad links between C and C'.

The experiments discussed in this chapter, which constitute a very small selection from a vast literature on consciousness research, are specific examples that reveal some aspect of the link between C and C'. Note that not everything in C' is mapped onto something in C. Which points in C' are mapped onto C? This is the underlying theme of most experiments we encountered in this chapter. One of the key findings of this line of research is the variable and tentative relationship between C and C', which is in a sense expected. Consciousness makes up its own contents even though there are counterparts to it in the external world, as was evidenced in the blind spot experiments. The individual variations in what is actually perceived by the subjects in blind spot and scotoma studies are another proof of this variable relationship between the stimulus and its conscious percept.

Another key observation that emerges from the consciousness studies of the kind described in this chapter is that there are conditions of adequacy that neural activity has to satisfy for it to enter conscious awareness. The adequacy could be in terms of intensity of stimuli, or the duration of stimulation as was seen in Libet's experiments. Or the condition could be in terms of patterning of neural activity, for example, the requirement of high synchrony in gamma range, for the neural activity to enter consciousness.

William James' insight suggests that consciousness is primarily an agent of selection, whether it is a selection of a conscious percept on the sensory side or the selection of an action on the motor side. We have seen such selection at play on the perceptual front in the binocular rivalry experiments. Both monkeys and human subjects did not see a simple superimposition of the rival patterns; they saw a single percept at any given time, though both stimuli were presented all the time separately to the two eyes. Each pattern could be represented by activity in an extended pool of neurons. It is possible that neuronal pools that represent the two patterns compete with each other by mutual inhibition, thereby simulating the selectional mechanism underlying conscious selection. Or such a selection could also be aided by feedback coming from higher visual cortical areas, or even from prefrontal areas. Binocular rivalry experiments also reveal that activity in V1 is only necessary but not a sufficient condition for rivalry to occur; when conscious rivalry occurs, V1 activity is usually accompanied by activity in higher areas. Similar findings come from word-masking studies by Dahaene and coworkers. When backward-masked visual words were presented to the subjects, the stimulus remained subliminal and activity was confined to the primary visual cortex. But when the subjects became conscious of words presented without the mask, there was widespread activation in visual, parietal, and frontal areas.

Thus, conscious percept seems to involve recruitment of a large number of brain areas; it is not a local confined activity, though it might begin as such. For neural activity to enter consciousness, it seems to be required to access what Bernard Baars calls the Global Workspace which integrates activities of small, local brain networks into an integrated whole. Activity, pertaining to a large number of specialized net-

works, which remains subconscious, forms part of conscious awareness when these specialized networks are integrated into a dynamic whole. Such integration seems to take place dynamically through synchronization.

In view of overwhelming evidence for such global integration supporting conscious activity, several schools of thought about consciousness seem to tend towards a common ground. Bernard Baars, a strong proponent of the Global Workspace theory, points out this convergence of several views of consciousness.

In an excellent book on consciousness, Gerald Edelman and G. Tononi state that, “when we become aware of something … it is as if, suddenly, many different parts of our brain were privy to information that was previously confined to some specialized subsystem. … the wide distribution of information is guaranteed mechanistically by thalamocortical and corticocortical reentry, which facilitates the interactions among distant regions of the brain.” [f] (pp. 148–149).

Edelman and Tononi have developed a theory according to which complex dynamics of a set of tightly coupled brain areas, collectively described as the *dynamic core*, is the prerequisite for conscious perception. The theory goes beyond the requirement of extensive recruitment of *structural* brain areas for conscious perception; it specifies *temporal* or dynamic conditions, in terms of precise complexity measures, on the activity of the dynamic core.

Walter Freeman at the University of Berkeley who did pioneering work on how odors are represented as stable chaotic states in the olfactory bulb of rabbits concurs that “the activity patterns that are formed by the (sensory) dynamics are spread out over large areas of cortex, not concentrated at points.”

Inspired by experimental data that points to multiregional synchronization as a requirement for conscious perception and extraction of meaning, neurologist Antonio Damasio expresses similar views. “Meaning is reached by time-locked multiregional retroactivation of widespread fragment records. Only the latter records can become contents of consciousness.”

The precise neural components that form part of the global workspace which serves as a stage for the play of conscious perception may vary from one theory to another. But the essential view that a global network is involved is the common feature of many of these theories.

Commenting on the crucial role of thalamus as a key hub in the brain, that links not only sensory channels with the cortex, but also various cortical areas with each other, Rodolfo Llinas expresses his ideas as follows: “… the thalamus represents a hub from which any site in the cortex can communicate with any other such site or sites. … temporal coincidence of specific and non-specific thalamic activity generates the functional states that characterize human cognition.”

The above theories of consciousness or approaches to consciousness, laudable they may be, considering the impressive convergence of a variety of approaches, do not represent the final word on the subject. They are correlational accounts of consciousness. They merely summarize the conditions under which neural activity of the brain enters consciousness. They cannot be treated as comprehensive mechanistic, or “physical” theories of consciousness.

A few physical theories of consciousness have indeed been proposed. Susan Pockett and Johnjoe McFadden have independently proposed the idea that the substrate for conscious operations in the brain is the electromagnetic field generated by neural electrical activity. Synchronous neural activity is likely to produce stronger fields, and stronger conscious correlates, which is also consistent with the earlier proposals that synchrony in gamma range is related to consciousness.

A comprehensive physical theory of consciousness would not only account the cognitive data in terms of correlations; it also explains *why* certain neural states are associated with consciousness. It would first and foremost define consciousness, quantify it and even present units of the same, just as a theory of any physical quantity would. The state of art seems to be quite far from that goal. To employ an analogy from the history of electricity and magnetism, contemporary theories of consciousness are at the stage of leyden jar and lodestone. The moment of Maxwell equations is not here as yet. The distortions in our perception of space and time discovered by consciousness researchers must be treated as real effects, with consequences to real-world space and time, and not just as anomalies of the mind, just as the “anomalies” in Michelson–Morley experiment, when taken seriously, led to creation of a revolutionary theory of space-time a century ago. Do we need to wait for another century for the emergence of a comprehensive and satisfactory theory of consciousness that is capable of marrying perfectly the subjective and objective worlds in a single, integrative theoretical framework?

## References

- Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences*, 6(1), 47–52.
- Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience? *Progress in Brain Research*, 150, 45–53.
- Bach-y-Rita, P. (1972). *Brain mechanisms in sensory substitution*. New York: Academic Press.
- Barker, A. T., Jalinous, R., & Freeston, I. L. (1985). Non-invasive magnetic stimulation of human motor cortex. *The Lancet*, 1(8437), 1106–1107.
- Blake, R., & Tong, F. (2008). Binocular rivalry. *Scholarpedia*, 3(12), 1578.
- Botvinick, M., & Cohen, J. (1998). Rubber hands “feel” touch that eyes see. *Nature*, 391, 19.
- Brindley, G. S., & Lewin, W. S. (1968). The sensations produced by electrical stimulation of the visual cortex. *The Journal of Physiology*, 196, 479–493.
- Crick, F. (1995). *The astonishing hypothesis: The scientific search for the soul, scribner reprint edition*. Simon & Schuster Adult Publishing Group.
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25–62.
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.-F., Poline, J.-B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4(7), 752.
- Dennet, D. C. (1993). *Consciousness explained*. London: Penguin Books.
- Descartes, R. (1641/1984). Meditations on first philosophy. In *The philosophical writings of René Descartes* (Vol. 2, pp. 1–62) (J. Cottingham, R. Stoothoff, & D. Murdoch, Trans.). Cambridge: Cambridge University Press.

- Edelman, G., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination.* New York: Basic Books.
- Freeman, W. J. (1991). The physiology of perception. *Scientific American*, 264, 78–85.
- 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, 1698–1702.
- Holt, J. (1991). *How children learn* (pp. 18–20). London: Penguin Publishers.
- James, W. (1890). *The principles of psychology* (p. 136). Cambridge: Harvard University Press. 1983 paperback, ISBN 0-674-70625-0 (combined edition, 1328 pages).
- Kornhuber, H. H., & Deecke, L. (1990). Readiness for movement—The Bereitschaftspotential-story. *Current Contents Life Sciences*, 33(4), 14.
- Lee, S.-H., Blake, R., & Heeger, D. J. (2005). Travelling waves of activity in primary visual cortex during binocular rivalry. *Nature Neuroscience*, 8(1), 22–23.
- Lee, S.-H., Blake, R., & Heeger, D. J. (2007). Hierarchy of cortical responses underlying binocular rivalry. *Nature Neuroscience*, 10(8), 1048–1054.
- Libet, B. (1965). Cortical activation in conscious and unconscious experience. *Perspectives in Biology and Medicine*, 9, 77–86.
- Libet, B. (1981). The experimental evidence for subjective referral of a sensory experience backwards in time: Reply to PS Churchland. *Philosophy of Science*, 48, 182–197.
- 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.
- Logothetis, N., & Schal, J. (1989). Neuronal correlates of subjective visual perception. *Science*, 245(4919), 761–763.
- McFadden, J. (2002). The conscious electromagnetic information (Cemi) field theory: The hard problem made easy? *Journal of Consciousness Studies*, 9(8), 45–60.
- Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W., & Rodriguez, E. (2007). Synchronization of neural activity across cortical areas correlates with conscious perception. *The Journal of Neuroscience*, 27(11), 2858–2865.
- Penfield, W. (1958). *The excitable cortex in conscious man*. Liverpool: Liverpool University Press.
- Pockett, S. (2000). *The nature of consciousness*. Writer Club Press, Lincoln, Nebraska.
- Porta, J. B. (1593). *De refractione. Optices parte. Libri novem*. Naples: Salviani.
- Ramachandran, V. S. (1992, May). Blind spots. *Scientific American*, pp. 86–91.
- Ramachandran, V. S., & Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350, 699–702.
- Schmaltz, T. (2002). Nicolas Malebranche. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2002 Edition). <http://plato.stanford.edu/archives/sum2002/entries/malebranche/>.
- Singer, W. (2007). Binding by synchrony. *Scholarpedia*, 2(12), 1657.
- Stapp, H. (1993). *Mind, matter and quantum mechanics*. Berlin: Springer.
- Tecchio, F., Babiloni, C., Zappasodi, F., Vecchio, F., Pizzella, V., Romani, G. L., et al. (2003). Gamma synchronization in human primary somatosensory cortex as revealed by somatosensory evoked neuromagnetic fields. *Brain Research*, 986(1–2), 63–70.
- Walter, W. G. (1963). *Presentation to the Osler society*. Oxford: Oxford University Press.
- Watson, J. B. quote: Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158–177.
- Weiskrantz, L. (1988). Some contributions of neuropsychology of vision and memory to the problem of consciousness. In A. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 183–199).
- Weiskrantz, L. (1997). *Consciousness lost and found. A neuropsychological exploration*. Oxford: Oxford University Press.
- Weiskrantz, L. (2007). Blindsight. *Scholarpedia*, 2(4), 3047.
- Wheatstone, C. (1838). Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128, 371–394.