

Andrea Eugenio Cavanna
Andrea Nani

Consciousness

Theories in Neuroscience
and Philosophy of Mind

 Springer

Consciousness

Andrea Eugenio Cavanna • Andrea Nani

Consciousness

Theories in Neuroscience
and Philosophy of Mind

 Springer

Andrea Eugenio Cavanna
Michael Trimble Neuropsychiatry
Research Group
BSMHFT and University of Birmingham
Aston University
Birmingham
UK

Andrea Nani
Michael Trimble Neuropsychiatry
Research Group
BSMHFT and University of Birmingham
Birmingham
UK

ISBN 978-3-662-44087-2 ISBN 978-3-662-44088-9 (eBook)
DOI 10.1007/978-3-662-44088-9
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014950510

© Springer-Verlag Berlin Heidelberg 2014

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

*To our parents
who taught us to appreciate consciousness
in all its niceties*

Contents

Part I Philosophical Theories of Consciousness

1	David Chalmers	3
2	Paul and Patricia Churchland	9
3	Tim Crane	15
4	Donald Davidson	19
5	Daniel Dennett	25
6	René Descartes	29
7	Jerry Fodor	37
8	Jaegwon Kim	43
9	William Lycan	49
10	Colin McGinn	55
11	Thomas Nagel	61
12	Alva Noë	67
13	Hilary Putnam	73
14	David Rosenthal	79
15	John Searle	85

Part II Scientific Theories of Consciousness

16	Bernard Baars	93
17	Francis Crick and Christof Koch	99
18	Antonio Damasio	105

19	Stanislas Dehaene	111
20	Merlin Donald	117
21	John Eccles and Karl Popper	123
22	Gerald Edelman	127
23	Nicholas Humphrey	133
24	Julian Jaynes	139
25	Benjamin Libet	145
26	John Kevin O'Regan	151
27	Roger Penrose and Stuart Hameroff	157
28	Giulio Tononi	163
29	Max Velmans	169
30	Semir Zeki	175
	Epilogue: A Brief Tour of the Introductions to Consciousness Studies . . .	181
	Bibliography	191
	Index	197

Introduction: The Conundrum of Consciousness

*The feeling of an unbridgeable gulf between consciousness and brain-process:
how does it come about that
this does not come into the considerations of our ordinary life?
This idea of a difference in kind is accompanied by slight giddiness,
which occurs when we are performing a piece of logical sleight-of-hand.
[...] When does this feeling occur in the present case?
It is when, I, for example, turn my attention in a particular way
on to my own consciousness,
and, astonished, say to myself:
THIS is supposed to be produced by a process in the brain!*
(Ludwig Wittgenstein, *Philosophical Investigations*, part I, section 412)

The study of the human mind and consciousness has never been so exciting. During the last few decades, both philosophical and scientific attempts to account for the origin and nature of consciousness have mushroomed so as to form a respectable field of research that has conquered autonomy and independence. The main goal of the present book is to provide the reader with an overview of what we consider to be the two faces of consciousness studies. On the one hand, philosophers of the mind have explored a number of different theoretical positions in order to unravel and disentangle the conceptual intricacies of consciousness. On the other hand, neuroscientists, psychologists, and cognitive scientists have developed a panoply of theoretical and empirical models of the brain mechanisms underlying consciousness.

The demarcation line between philosophers and scientists is far from straight and univocal; however, it rests upon the observation that philosophers and scientists tend to adopt different perspectives in their approach to the study of consciousness, mind, and brain. The philosophical approach traditionally devotes more attention to logical reasoning and aims to develop and describe the most coherent pictures of different possible scenarios. Therefore, philosophers are mainly concerned with concepts; their purpose is to examine and sift out different conceptual possibilities in order to provide a theoretical framework for a valid account of mental phenomena. To do so, philosophers can use a set of conceptual tools, including the often-employed recourse to modal logic. Thought experiments (*Gedankenexperiment* in the philosophy of mind jargon) can be particularly helpful as they can offer theoretical proofs in order to foster the plausibility of a speculative theory over competitor ones. Well-known thought experiments in the philosophy of mind are, for instance, the conceivability of doppelgangers and of inverted qualia, Mary the color-blind neuroscientist, the Chinese room, the philosophical zombies, etc. Provided we

accept certain premises, thought experiments ask us to conceive possible worlds and counterfactual scenarios, so that we can derive conceptual consequences which help to evaluate the logical solidity of theoretical positions.

Contrary to philosophers, most neuroscientists (including experimental psychologists) are less familiar with thought experiments and modal logic. They are more inclined to the experimental applications of scientific techniques which can provide useful insights into the brain mechanisms which bring about our mental functions. Scientists are also deeply interested in giving the most correct interpretation of the amount of data collected in laboratory experiments, which can eventually lead to a coherent account of empirical evidences. Contemporary neuroscientists marshal investigations using structural and functional imaging techniques (e.g., PET and fMRI scans of living brains) and observations of neurological cases which can dovetail observable behaviors and brain functions with mental phenomena as simple as, say, visual percepts. The hypothesis at the basis of these investigations is that the activity of the brain can explain every aspect of our mental life, namely, that brain processes can actually provide the essential substrates for perceptions, sensations, attention, consciousness, and all the other attributes that compound the human mind. According to the neuroscientific approach, every mental state – such as being conscious of the chocolate taste – should be realized by specific brain activities: the so-called neural correlates of consciousness.

At first glance, the philosophical and the scientific strategies to tackle the mind-body problem and the puzzle of consciousness seem very different. The former analyzes concepts and makes them clearer by screening out logically implausible concepts; the latter investigates specific mental functions in order to identify the cerebral areas and brain processes involved in their realization. It would however be a mistake to simplistically consider philosophy as a purely conceptual task and science as a merely empirical enterprise. Science cannot be deaf to conceptual analyses, as much as philosophy cannot be blind to scientific results. Arguably, the most fruitful methodology for studying the human mind is to favor a multidisciplinary and highly integrated approach, capable of combining relevant findings from different disciplines. This is the rationale for the dual nature of this book.

The first section is dedicated to the philosophical endeavor of understanding the conscious mind. The philosophers presented in this book summarize a wide spectrum of theoretical positions in the philosophy of mind. In a sense, they can be seen as the champions of distinct and clear philosophical models on human consciousness.

David Chalmers has purported an interesting form of dualism between the mental and the physical, called “property dualism.” According to this view, the mental aspect is seen as an irreducible and fundamental characteristic of matter, together with other characteristics that are merely physical.

Paul and Patricia Churchland have always maintained an intransigent position with regard to the mental, providing the theoretical background for a future neuroscience which is deprived of any account given in the first-person perspective. Their position is known as “eliminative materialism.”

Tim Crane has convincingly argued in favor of considering intentionality as the hallmark of mental phenomena. According to his view, even though the mind is necessarily brought about by brain functions, mental contents cannot be reduced to neurophysiology. In fact, they appear to be ontological entities of a new kind.

Donald Davidson was a philosopher of the mind who famously argued in favor of the “token identity theory” between mental states and physical brain states, while claiming at the same time the impossibility of psychophysical laws capable of giving a nomological reduction of the mind to the brain. Hence, he called his doctrine “anomalous monism.”

Daniel Dennett has inaugurated a new method for studying the mind and the origin of consciousness. He names his approach of examining mental phenomena “heterophenomenology” and develops what he refers to as the “multiple drafts model” theory of consciousness.

René Descartes is commonly considered the father of the modern philosophy of mind, since he developed the first theoretical position on this subject (classical dualism), which holds that mind and matter are different substances. Classical dualism has proved to be a thought-provoking and never-ending source of debates (especially in modern neurosciences: see, for example, Damasio’s discussion of Descartes’ philosophy in his *Descartes’ Error* and Libet’s original interpretation of Descartes’ dualism in his *Mind Time*).

Jerry Fodor is the proponent of philosophical theories (language of thought and modularity of mind) which have gained remarkable interest within both the communities of philosophers and cognitive scientists. His acknowledged contributions represent a painstaking analysis of concepts which form our ordinary way of speaking about mental processes.

Jaegwon Kim is one of the most inspired contemporary philosophers of the mind. His theoretical approach to the mind-body problem is based on the investigation of the concept of “supervenience,” or how mental phenomena supervene from the specific organization of neuronal patterns. Kim also provides a thorough analysis of mental causation and its consequences for the physicalist doctrine.

William Lycan provides a specific line of attack against the theoretical difficulties that characterize the analysis of the nature of consciousness. He argues that a cluster of different issues constitutes the problem of consciousness and therefore proposes to carefully divide this problem into separate issues which can be tackled more easily one by one. In Lycan’s view, all the features of the conscious mind can be exhaustively accounted for by virtue of its representational properties and of the functional organization of its components.

Colin McGinn holds an original and controversial position within the philosophy of mind, as he argues that the intrinsic nature of consciousness might be characterized by nonspatial features. This elusive nature is responsible for preventing us to really understand conscious phenomena. He thus claims that consciousness is destined to remain a mysterious aspect of reality, which we cannot fully comprehend and, therefore, explain.

Thomas Nagel is the author of one of the most famous articles in the field of philosophy of mind (“What is it like to be a bat?”), in which he argued against the reductionist version of physicalism. Nagel’s argument is that reductionist physicalism cannot give a satisfactory account of the phenomenal quality of mental phenomena (in other words, “what it is like to be that mind”). He also develops an intriguing theoretical position, based upon a non-reductive approach that goes beyond the division between subjective and objective perspectives to show that they are inextricably connected, in what he refers to as “the psychophysical nexus.”

Alva Noë argues that conscious experience results from a skillful activity of the body, which in turn depends on the interaction among the body, the brain, and the environment. This thought-provoking “enactive” approach contends the common scientific assumption that consciousness is a functional activity confined within the brain and argues in favor of a broader perspective which conceives the conscious perceptual process as active engagement with the world.

Hilary Putnam wrote fundamental essays on the multiple realization of mental properties throughout functional physical processes. However, in his later writings, he refutes functionalism and highlights the pragmatic importance of mental states for the conscious behavior of human beings.

David Rosenthal is a prominent exponent of the higher-order theory of consciousness, which is based on the idea that conscious states arise as a result of specific thought processes. Consciousness is therefore conceived of as a further level in the elaboration of information, since thoughts about mental states are on a higher order with respect to their contents.

John Searle is a leading figure in the philosophical debate concerning mind and consciousness, and his profile concludes the first section of this book. In his works, he makes it clear that he is neither a materialist/eliminativist with regard to the mental domain nor a dualist who accepts the mind as a different substance opposed to the physical world. According to his view (called “biological naturalism”), we should see any mental state as a biological phenomenon, such as digestion and the production of bile. Since the mind finds its place in nature, all we need is a naturalistic explanation of how the brain works.

The second section is dedicated to the scientific theories of consciousness. Similarly to the philosophers presented in the first section of this book, the neuroscientists whose theories are presented here adopt a wide-ranging scale of different perspectives on the scientific study of mind and consciousness.

Over 20 years ago, cognitive scientist Bernard Baars put forward a theory of consciousness which constitutes one of the most reliable frameworks for fruitful research programs. Baars’ core idea is that consciousness is brought about by cerebral processes in a global workspace instantiated by certain areas of the brain. This kind of global workspace or “theater” is supposed to give a unified and coherent representation of perceptions, sensations, and thoughts.

Francis Crick and Christof Koch made fruitful collaborations in the scientific study of mind and consciousness. The late Francis Crick was among the first scientists to take consciousness seriously as an object of scientific inquiry. He suggested that consciousness could emerge from a uniform pattern of neural activity. Christof

Koch is following the line of work opened by Crick and is actively engaged in the quest for the neural correlates of consciousness, i.e., the biological hallmarks that specific conscious processing (which can result in visual or auditory experiences, for example) is actually going on in the brain.

Antonio Damasio is a leading neurologist who proposes a hierarchical theory of consciousness largely based on his observations on patients with neurological disorders. First, he postulates the existence of a basic form of consciousness, which is generated by neuronal configurations capable of comparing different states of the organism with regard to internal and external stimuli. Memory and the temporal lapse among the neural events play a crucial role in this comparison, because they allow to distinguish between first- and second-order cerebral representations. Damasio also identifies an extensive form of consciousness with wide temporal boundaries, based on the preceding brain mechanisms. This last form of consciousness is, in turn, fundamental for the construction of an autobiographical self.

The global workspace model originally proposed by Bernard Baars has been further developed by neuroscientist Stanislas Dehaene. Dehaene's model refines Baars' idea that consciousness is a mode of sharing information within a global workspace (in the case of the brain, *global neuronal workspace*) and identifies the fundamental signatures that can indicate whether or not the brain is conscious at any given time.

Merlin Donald is a psychologist who has focused his research on the problem of the origin of consciousness in the development of language, culture, and society. From this perspective, consciousness is seen as a cognitive device capable of prompting a long-range system for the guidance of our behavior. Donald offers a multilevel theory of our conscious capacities recognizing three main stages in the conscious activity of human beings. At the first level, consciousness operates in order to build a unified scenario from different perceptions. At the second level, consciousness can expand sensations into a short-term working memory space. At the third level, consciousness acts as an extended awareness of multifarious episodic representations, which leads to the formation of a fully fledged symbolic mind.

Neuroscientist John Eccles and philosopher Karl Popper coauthored the challenging book *The Self and Its Brain*, which portrayed interactionism as the most fruitful theoretical approach to tackle the mind-body problem. The kernel of their theory is to single out three autonomous worlds, the first one populated by physical objects, the second one by mental objects, and the third one by conceptual objects. According to Popper, each world is supposed to interact with the two others in dynamic processes that spring up our experience. Eccles presents a wide review of clinical cases and experimental results in order to build the interactionist thesis on solid neuroscientific grounds. Furthermore, he advances the hypothesis that the unity of our experience might be provided by an essential ontological structure, which he dubs the "self-conscious mind."

Gerald Edelman is regarded as a leading scholar in neuroscience. In his writings, he examines the cerebral architecture at both microscopical and macroscopical levels in order to figure out the way the brain can evolve. He argues that different configurations, maps, or patterns of neurons compete with each other to gain constancy and stability within the brain, an approach which has become famous as

“neural Darwinism.” In light of that, neuronal patterns are subject to natural selection processes, similarly to biological organisms whose species are selected by the environmental conditions. Edelman makes use of both neuroscientific and philosophical skills in order to give an enthralling account of the mind and its place in nature, by endorsing a sophisticated epiphenomenal position with regard to consciousness.

Theoretical psychologist Nicholas Humphrey suggests an original use of the concept of *intentionality* within the neuroscientific framework. He also questions our ordinary way of speaking about mental phenomena and proposes a new term (*sentition*) in order to indicate when a subject is consciously involved in a perceptual activity. According to Humphrey, consciousness has a specific and important role in the evolutionary development of *Homo sapiens*, as conscious experience is fundamental to our appreciation of life.

Julian Jaynes was a gifted psychologist who dedicated his efforts to the problem of the origin of consciousness. In a popular and thought-provoking book titled *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, he proposes the astonishing hypothesis that consciousness – defined as a higher-order form of self-conscious representation – arose out of the breakdown of a primitive bicameral mind devoid of truly subjective experiences. He supports his theory with insightful examples taken from ancient literatures – especially from the *Iliad* – in order to show that archaic men lacked our modern concept of consciousness.

Benjamin Libet was a neuroscientist who linked his name to a famous series of experiments showing that there is a measurable time delay between the decision to initiate an action and the conscious intention to do it. The interpretation of the results of Libet’s experiments is still a source of attention-grabbing debates, because of its bewildering implications for the problem of free will. Libet is also the author of a fascinating hypothesis for the nature of consciousness, which reinterprets the Cartesian dualism in a strikingly original way. According to his theoretical standpoint, consciousness is produced globally by the neocortex as a mental conscious field.

Kevin O’Regan is one of the strongest advocates of the intriguing sensorimotor approach to the nature of the conscious mind. According to this view, consciousness results from an activity during which the body and the environment are strictly intertwined. This model broadens the scope of neuroscience as it claims that consciousness is more than a mere property of the brain.

Roger Penrose is a leading mathematician who achieved important results both in mathematics and cosmology. His contribution to the field of consciousness studies is the theory that conscious states derive from non-computational operations based on quantum coherence in subcellular structures such as microtubules or nanotubes. Anesthesiologist Stuart Hameroff has joined Penrose in a multidisciplinary attempt to develop a consistent quantum mechanical approach to the study of consciousness.

Giulio Tononi has put forward one of the most interesting and promising theories of consciousness of the recent years. He argues that consciousness derives from a

specific mode of elaboration of information, which must be, at the same time, both integrated and differentiated. Further development and refinement of this theory could allow an accurate measurement of the degree of consciousness not only in humans but also in other species.

Max Velmans is a theoretical psychologist who emphasizes the ontological relevance of the first-person perspectives in order to get a complete account of mind and consciousness. Specifically, he claims that the common distinction between objective and subjective accounts does not make scientific sense. He has therefore developed a theory called “reflexive monism,” in which the phenomenal character of consciousness is to be conceived in a much broader sense, so that its contents expand to include the whole perceptual world.

Neurologist Semir Zeki, the last author portrayed in this book, has proposed a theory of multiple consciousnesses that are distributed in time and space. He therefore challenges the shared experience of a unitary and global conscious experience. Zeki also claims that the unity of consciousness has been overemphasized by both philosophers and neuroscientists, and this could overlook its real composite nature.

Our list of philosophers and scientists of the mind is far from being exhaustive, reflecting a personal pathway rather than a consensus-based selection of theories. However, it is worth noting that all the authors presented in this book are still active or have been so until recently. The only exception is Descartes, who set the theoretical framework for the modern approaches to the understanding of consciousness. Overall, we have tried to provide the reader with the widest overview of both philosophical and scientific theories of the conscious mind which are currently under discussion. The list is by necessity incomplete, and we are aware that important thinkers and researchers could not be included in the present edition. Each chapter has been dedicated to an individual author and, therefore, can be read independently of the others. We hope that this structure can facilitate a better understanding of the variety of interesting positions in the contemporary debate. However, although both philosophical and scientific chapters can be thought of as autonomous, certain themes are recurrent and interconnected, as we have tried to highlight with cross-references between several chapters.

In a sense, this book is a gallery of other books, just as a painting showing a gallery of containing other paintings (Fig. 1). In our attempt to provide a sort of compass that the readers can use as a guide for their exploration within the field of consciousness studies, we have tried to clarify and simplify a complex material. In doing so, we might have committed multiple sins from oversimplifications to misinterpretations and misvaluations. We can only apologize for any involuntary distortion of the theories of both philosophers and scientists, hoping that the benevolent readers will understand the nature of our effort and forgive any possible overlooking or inaccuracy.

As Wittgenstein reminds us, nothing seems to be so near and intimate and, at the same time, so remote and inaccessible from us than our own conscious mind. This could be defined as the *paradox of consciousness*: although everyone intuitively



Fig. 1 David Teniers the Younger (1610–1690), *The Archduke Leopold Wilhelm in His Gallery* (circa 1650–1652), oil on canvas, Kunsthistorisches Museum, Vienna, Austria

knows what it means to be conscious, there is still the astonishing impression of an unbridgeable gulf between our inner conscious world and what is out there. The theories presented in this book tell the story of how the human mind audaciously tries to build a walkable bridge over this gulf.

Birmingham, UK
Birmingham, UK

Andrea Eugenio Cavanna
Andrea Nani

Part I

Philosophical Theories of Consciousness

Property Dualism

*The really hard problem of consciousness is the problem of experience.
[...] Why should physical processing give rise to a rich inner life at all?
It seems objectively unreasonable that it should, and yet it does.
If any problem qualifies as the problem of consciousness, it is this one.*
(Facing up to the problem of consciousness)

Australian philosopher David Chalmers (born on 20 April 1966 in Sydney) stands out among the thinkers who have mostly contributed to revitalize the field of philosophy of mind over the last few years. His 1996 book *The Conscious Mind* had the ambition of setting the philosophical agenda with regard to the investigation of consciousness. Chalmers' philosophical interests also cover metaphysics and epistemology (branches of philosophy that deal with the nature of existence and knowledge), in which he has given relevant contributions with thought-provoking ideas. Chalmers gained his PhD from Indiana University Bloomington under Douglas Hofstadter. From 2002 to 2004, he was Director of the Center for Consciousness Studies at the University of Arizona and sponsor of the *Toward a Science of Consciousness* conference. He currently is Distinguished Professor of Philosophy and Director of the Centre for Consciousness at the Australian National University, and Professor of Philosophy and Co-director of the Center for Mind, Brain, and Consciousness at New York University.

Chalmers neatly demarcates the study of consciousness between an easy problem and a hard problem and claims that reductive methods are inadequate to solve the hard problem. Easy problems are instead those problems that are susceptible to be tackled with the standard methods of cognitive science and to be accounted for in terms of computational or neural mechanisms. Examples of easy problems are the following:

- The ability to distinguish, classify, and react to external stimuli
- The integration of different streams of information
- The reportability of mental states
- The accessibility of mental states
- The faculty of attention
- The deliberate control of behavior
- The difference between sleep and wakefulness

Although the abovementioned phenomena are all associated with consciousness, Chalmers argues that neuroscience is already well equipped to investigate their nature and sooner or later will be able to thoroughly understand how the brain can produce them.

By contrast, Chalmers claims that the really hard problem of consciousness is related to experiential contents, in particular to the subjective aspect of conscious experience. In fact, at any moment, we are experiencing subjective feelings, emotions, streams of thoughts, and perceptual sensations, such as the blueness of the sky, the softness of wool, the patter of rain on the roof, etc. All these experiences, coming from different modalities of perception and interoception, are united in a single conscious state. How the brain can bind all these different contents (the so-called binding problem) and *what it is like to be* this unified conscious state constitute, according to Chalmers, the hard problem of consciousness (about “what it is like to be something,” see also Chap. 11). But how is it possible that physical processes going on inside the brain could give rise to the richness of consciousness? Chalmers holds that neuroscience can offer us only quantitative descriptions of what the brain does, but it seems that there is an incommensurable gap between the detailed quantitative accounts given by neurophysiology and the qualitative aspects of our inner lives (“explanatory gap”).

As the qualitative or phenomenal features of consciousness are what make us human beings, a complete theory of consciousness should be able to address and solve the hard problem. Chalmers claims that phenomenal consciousness does not concern either a cognitive ability or a function, hence his belief that the methods of cognitive science are ill-suited for an explanation of consciousness. We can in fact describe the mechanisms by which an ability or a function is performed and the causal role that such ability or function plays in producing the resulting behavior. However, when we deal with consciousness, even though we can describe all the cognitive, behavioral, and neural underpinnings during the occurrence of a conscious experience (such as perceptual discrimination, categorization, internal access, verbal report, patterns of neuronal firings, etc.), there always remains an unanswered question: Why are these mechanisms specifically accompanied by conscious experience? According to Chalmers, a simple explanation in physical terms leaves this question totally open.

A materialistic reply to Chalmers’ line of reasoning is that neuronal processes, patterns, and mechanisms in the brain are not *accompanied by* conscious experience, but *are* themselves the very conscious experience. Chalmers adopts a modal argument based on the conceivability of philosophical zombies against this identity theory between mind and brain, which equates physical processes to mental phenomena. Since it is possible to conceive the idea of a zombie, that is, a perfect physical copy of a person but with no inner conscious experience, then the relationship between the conscious mind and brain is to be contingent rather than necessary. This mental experiment is used to refute the identity theory, because if this theory were correct, the relationship between mind and brain should be taken as necessary rather than simply contingent.

Some philosophers think that the modal argument is valid and able to discard the identity claim between mind and brain. Other philosophers, by contrast, consider



Fig. 1.1 Death of the automaton Talos, the horsemen holding Talos are the Dioscuri, Castor, and Pollux. Wilhelm Heinrich Roscher (1845–1923), *Ausführliches Lexikon der griechischen und römischen Mythologie*, 1884

this argument as fallacious and misleading. Their main point is that the argument is logically incoherent, because if mind and brain are the same entity, then it is not coherently possible to imagine one without the other. In other words, following the philosophical zombie argument would be similar to conceiving a pencil that does not write. What kind of pencil would it be? And would it still be a pencil?

Whatever may be the case, it seems to us that along the philosophical zombie argument it would be possible to conceive a complementary argument, in which a perfect copy of a person's consciousness is insufflated into an artificial body. We could call this animated body an *automaton* and this mental experiment the *argument of automaton*, respectively, in analogy to the ancient Greek myth of automata, which were thought to be metallic statues of animals, men, and monsters crafted and made alive by the divine smith Hephaestus (Fig. 1.1). Needless to say, the conceivability of automata seems to be much more counterintuitive than the conceivability of philosophical zombies, but if you accept the latter, you must also accept the former.

The philosophical zombie argument necessarily leads to a dualistic conclusion: conscious mental events and physical processes are not the same things, because according to the perspective of physics, everything in the world seems to be compatible with the absence of consciousness. Thus, although the conscious mind *arises* from brain activity, this does not imply that brain activity *entails* consciousness. However, Chalmers' dualistic perspective does not consider mind and matter as two different and separate substances, as it is held in classic Cartesian dualism

(see Chap. 6), but instead suggests to take *experience* as a new fundamental property of reality. In this view, experience would be the additional ingredient that is not considered in the physical accounts of consciousness, an ingredient that is nonphysical in its nature and cannot be further reduced to more essential elements.

Chalmers therefore proposes a non-reductive explanation of consciousness and calls this approach *naturalistic dualism* or *property dualism*. According to this doctrine, experience is to be thought of as a fundamental element of the world, alongside mass, charge, and space-time. Chalmers considers this position an innocuous version of dualism, because nothing within it seems to contradict the physical knowledge we have of the world. As the approach postulates a new basic property of reality, it requires the addition of “bridging laws” in order to explain how experience can arise from the physical substrate. In other words, a non-reductive theory of consciousness is to be based on psychophysical principles, which are able to connect the properties of physical processes to the properties of experience. These principles should reveal what kind of physical systems can be associated with experience, what kinds of physical properties are significant for the emergence of experience, and what types of experiences a specific system can realize.

Chalmers puts forward three principles: structural coherence, organizational invariance, and a double-aspect view of information.

The principle of structural coherence states a direct correspondence between the *structure of consciousness* and the *structure of awareness*. Rather than a synonym of consciousness, for Chalmers, “awareness” refers to a specific process in the cognitive organization of experience; in particular, awareness should be thought of as *direct availability of information for global control*. Thus, whenever there is conscious experience, there is also awareness, that is, direct availability of certain information to the cognitive system for the control of behavior and verbal report. Similarly, whenever information is available for global control and verbal report, there is also a corresponding conscious experience.

The principle of organizational invariance states that if two systems have the same fine-grained *functional organization*, they will also have qualitatively identical experiences. Accordingly, what matters for the emergence of conscious experience is not the physical underpinnings of the system (such as the material nature of its components), but rather the abstract shape of causal interactions occurring between its components. Thus, this principle predicts that computers will be conscious, when they are able to replicate the functional organization of the human brain. As we shall see in the next chapters, many philosophers maintain that consciousness merely depends on a specific functional organization, which in turn does not rely on the characteristic structure of the system. However, although in some cases it is true that the same function can be perfectly replicated using different materials (think about a bowl made of glass, wood, or gold), this rule could not apply to all functions. With regard to consciousness, neuroscience is revealing that the anatomico-physiological properties of neurons play a fundamental role in causing conscious processes, and these specific properties cannot be exactly replicated by silicon chips (for a similar criticism, see Chap. 5).

The last and most important principle proposed by Chalmers is the double-aspect theory of information. According to this principle, information has two features, one which is physical and another one which is phenomenal. Experience would arise, therefore, by virtue of being one of the two fundamental aspects of information. Information could be physically embodied within a space of distinct physical states, which in turn could be associated with experiential or phenomenal states, in analogy to the correspondence that can be found between the structures of consciousness and awareness. Thus, differences in phenomenal states would structurally correspond to differences in physical processes, so that the same information can be embedded in both a physical process and a conscious experience.

Among the three principles, the last one is the most important and basic, as it confronts us with a new metaphysical viewpoint on nature. In theory, all information might have an intrinsic phenomenal aspect, so that where there is simple information processing, there is also simple conscious experience, and where there is complex information processing, there also is complex conscious experience. And since information is everywhere, we should conclude that consciousness too is in different degrees everywhere. This position is called *panpsychism* by philosophers and considered by many as a very counterintuitive conception of reality, as it is difficult to believe that a thermostat may have a phenomenal experience whatsoever.

Essential Bibliography

- Chalmers D (1995) Facing up to the problem of consciousness. *J Conscious Stud* 2:200–219.

In this article Chalmers explains why he thinks that the study of consciousness implies the resolution of an easy problem on one hand, and a hard problem on the other. He then sketches the framework for a theory of consciousness able to address the hard problem, based on the three principles of structural coherence, organizational invariance, and a double-aspect theory of information.

- Chalmers D (1996) *The conscious mind: in search of a fundamental theory*. Oxford University Press, New York.

This book extensively presents with a clear and fascinating style Chalmers' considerations about the problem and the nature of consciousness. The philosophical argumentation develops in rigorous conceptual analyses with large use of thought experiments leading to challenging ideas.

- Chalmers D (2010) *The character of consciousness*. Oxford University Press, New York.

This book collects Chalmers' papers on consciousness from the publication of *The Conscious Mind* to the present. Chalmers discusses both philosophical and scientific approaches to the study of consciousness and further develops his thought-provoking ideas.

Neurophilosophy and Eliminative Materialism

*Bit by experimental bit, neuroscience
is morphing our conception of what we are.
The weight of evidence now implies
that it is the brain, rather than some nonphysical stuff
that feels, thinks, and decides.*
(Brain-Wise: Studies in Neurophilosophy)

Paul Churchland (born on 21 October 1942 in Vancouver, Canada) and Patricia Smith Churchland (born on 16 July 1943 in Oliver, British Columbia, Canada) are Canadian-American philosophers whose work has focused on integrating the disciplines of philosophy of mind and neuroscience in a new approach that has been called *neurophilosophy*. They are also central figures in the philosophical stance known as *eliminative materialism*.

In 1969, Paul Churchland gained his PhD from the University of Pittsburgh under the supervision of philosopher Wilfrid Sellars. He then took positions at the University of Toronto, at the University of Manitoba, and at the Institute of Advanced Study in Princeton. In 1983, he joined the University of California, San Diego, where currently is Professor Emeritus and Valtz Chair of Philosophy.

Patricia Smith Churchland studied at the University of Pittsburgh and the University of Oxford, from which she received her M.A. and B.Phil., respectively. From 1969 to 1982, she was Assistant Professor at the University of Manitoba; she then spent a year as Visiting Member at the Institute of Advanced Study in Princeton and another year as Full Professor at the University of Manitoba. She is currently UC President's Professor of Philosophy Emerita at the University of California, San Diego, where she has taught since 1984. She has also held an adjunct professorship at the Salk Institute for Biological Studies since 1989.

The Churchlands have supported the thesis that the commonsense conception of psychological states is to be thought of as an incorrect theory about both the function and nature of mind. In fact, they claim that the explanation of human behavior in terms of mental phenomena commits us to an ontology that is not sustained by an accomplished neuroscientific model of the mind. Folk psychology – that is, the psychology of our commonsense and ordinary way of speaking – leads us to believe in the existence of desires, beliefs, fears, intentions, and so on, but such mental states, the Churchlands hold, do not exist as real phenomena. Instead, mental states are similar to phlogiston, the substance that chemists once supposed to be released by bodies during combustion. And similarly to the theory of phlogiston, which was

considered as false by a more accomplished chemistry, folk psychology will be considered as an erroneous theory of mind by a matured neuroscience.

The Churchlands claim that when we are speaking of mental phenomena, we are in fact speaking of neurophysiological states, that is, patterns of neuronal firings, as well as functional and computational processes. In this view, the mind is no more and no less than what the brain does, so that neuroscientists will be able to eventually replace the classification and causal generalizations of folk psychology with another classification given in physical terms only. This will prove that there is just one real kind of ontology in the world, the physical one, which is assumed by neuroscience. Similar theoretical reductions or, in this case, *eliminations*, have already occurred several times in science (e.g., the identification of “temperature” with “mean molecular kinetic energy”), and there is good reason for thinking that the same thing will occur for mental states as well.

According to the Churchlands, at least four arguments can be put forward to claim the identity between mental phenomena and physical brain processes. First, both the constitution and origin of human beings seem to be manifestly physical. In other words, humans seem to be purely physical systems whose development is coded and programmed in DNA molecules within the cell nuclei. Second, the origin of each type of animal appears to have a physical nature. Moreover, the theory of evolution and natural selection maintains continuity between all the living organisms and provides sound reasons for considering the nervous system as the fundamental cause of human behavior. Third, there is overwhelming evidence in favor of the *neural dependence* of mental states. In fact, all mental phenomena appear to be variously affected by brain damage, and this is precisely what we should expect if the identity theory holds true. Fourth, neurosciences are progressing fast and encouragingly gathering findings that show how human behavioral capacities and deficits can be accounted for in terms of specific physical processes occurring in brain structures. In light of these arguments, the Churchlands claim that the neuroscientific endeavor provides a promising framework within which the identity between mind and brain will be finally proven.

Still, opponents of eliminative materialism claim that the reductionist approach of neuroscience cannot account for the phenomenal features of subjective conscious experiences. In other words, any neurobiological theory of consciousness will always leave out something crucial, the feeling of what it is like to be in a certain state of mind, for instance, to be aware of the smell of mint, and so on (see Chaps. 1 and 12). In contrast to this position, the Churchlands argue that the objection rests on a misunderstanding. In fact, the argument presumes that if a conscious experience (say, feeling the smell of mint) were finally explained by neuroscience, then anyone who can understand the neuroscientific account of how the brain feels the smell of mint should be led to have that very experience. But why, the Churchlands point out, should the understanding of a theory result in the production of the phenomenon that the theory explains? This would be to ask too much of not only neuroscience but of every other kind of theoretical framework.

Another objection commonly raised against the neurobiological understanding of mind and consciousness is that even if neuroscience were able to precisely

identify the brain processes involved in certain conscious experiences, we would still not comprehend why these particular brain processes and not others are identical with those conscious experiences. Why should that specific activation of the olfactory cortex correspond to feeling the smell of mint and not the fragrance of lavender? As we have seen in the previous chapter (and we shall see again in due course), philosophers make considerable use of thought experiments in which they can conceive of the same brain processes either as being deprived of qualitative features (e.g., the zombie argument) or as being endowed with many different qualitative attributes (e.g., the inverted spectrum, see Chap. 9, and the Twin Earth thought experiment, see Chap. 13). However, the Churchlands argue that this objection too rests on a misunderstanding. None of the disciplines of science can explain *why* a certain phenomenon is identical to another, because this sort of identities cannot be explained but simply discovered. Science has discovered that “temperature” is equivalent to “mean molecular kinetic energy” and that “light” is actually a type of “electromagnetic radiation,” and in these cases, no one would demand that science be able to find something more in order to explain why two descriptions refer to one and the same natural phenomenon. Science does not tell us why these equivalences occur, because they are just the way things are. There is no fundamental set of laws from which we can derive that temperature is mean molecular kinetic energy or that light is electromagnetic radiation. In our world, it just happens to be so. When we realized that the Morning Star (the planet Venus) is identical to the Evening Star (the planet Venus) – which are allegorically represented by the same girl in a lithograph by art nouveau artist Alphonse Mucha (Fig. 2.1) – there would be no appropriate answer to the question “why is the Morning Star (the planet Venus) identical to the Evening Star (the planet Venus)?” because that is just the way our world is.

The identity between mental and brain states happens to be exactly the same. If science discovered that a certain pattern of neuronal activity is identical to the conscious sensation of smelling lavender, then there would be no further need for explaining why *that* specific pattern of neuronal activity is identical to the conscious sensation of smelling lavender. It just is.

Thus, the Churchlands support a strict neurobiological approach to the study of consciousness, even though for its complexity the study of consciousness lies at the crossroads of different disciplines, such as neuroscience, psychology, philosophy, anesthesiology, genetics, ethology, and evolutionary biology. All these perspective, however, should be based on the common ground that mental processes such as thinking, feeling, and consciously experiencing are entirely physical brain events. Therefore, the classical mind-body problem should be reformulated as a mind-brain problem, that is, as a nest of empirical questions concerning the function of the brain. Within this neurobiological framework, typical questions that form the mind-brain problem are, for instance, understanding the functional difference in the brain between being awake or being in deep sleep, as well as understanding how these conditions can be compared to the loss of consciousness during absence seizures. Other important issues are understanding how much of the decision-making process is conscious and therefore what the differences are between conscious and unconscious stages of decision-making; what the exact relationship is between

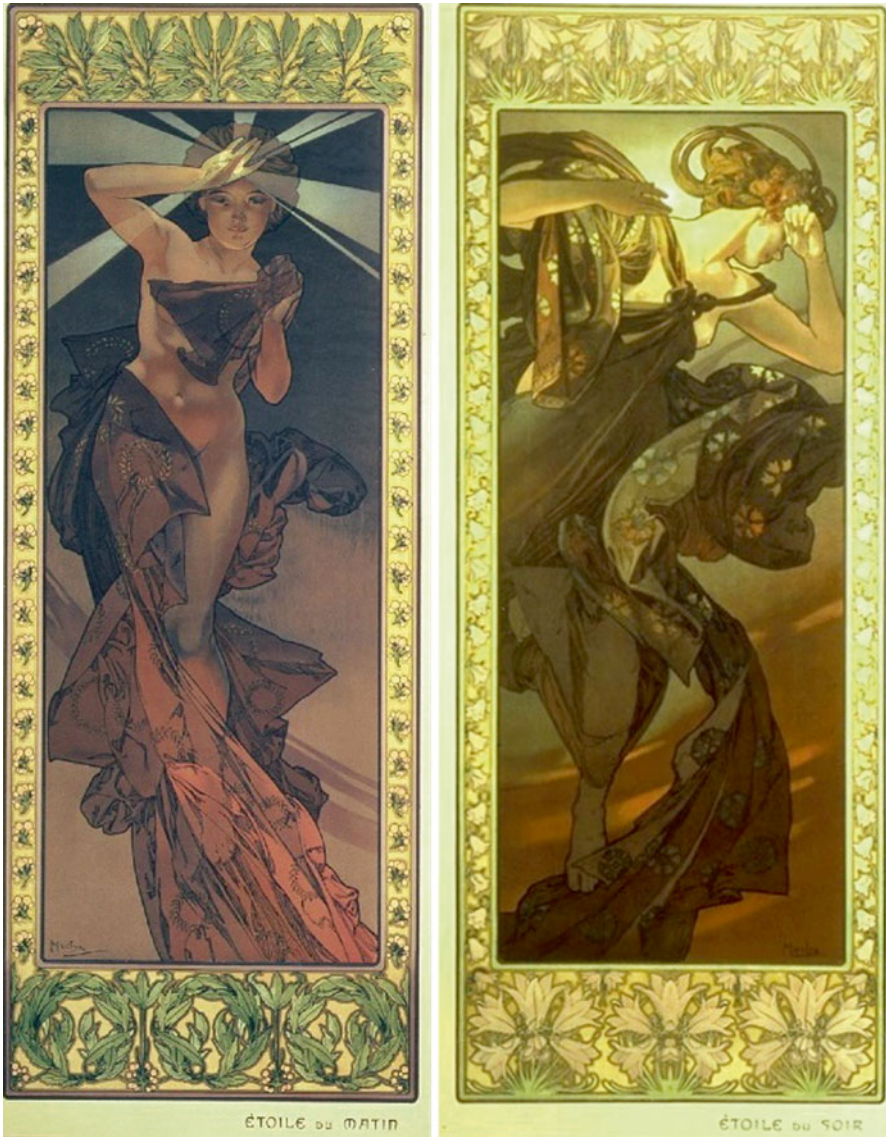


Fig. 2.1 Alphonse Mucha (1860–1939), *Morning and Evening Star*, cropped image from the art poster *The Moon and the Stars* (1902)

consciousness and attention and what happens in the brain when new skills and movements, which were acquired consciously, can be performed automatically.

Consciousness appears to be related to different kinds of phenomena, and one of the tasks of the neurophilosopher is, according to the Churchlands, trying to unravel the conceptual intricacies of neuroscientific discoveries. Neurophilosophers mainly deal with the nature of mind, representation, rationality, knowledge, and morality by

integrating philosophical analyses with neuroscientific findings. Neurophilosophy is at present an expanding and promising field of research. Within it, Paul Churchland is currently more active in developing a neurobiological approach of how the brain can acquire knowledge, whereas Patricia Smith Churchland is more interested in discussing how neuroscience has radically changed the study of human morality and free will.

The Churchlands have contributed with original ideas to some classical themes of philosophy by suggesting new perspectives and insights. Still, neurophilosophy should not be thought of as the canonical interpretation of neuroscience but just one of many ways of reading neuroscientific results. The exploration of the brain is at its beginning, and no one can foresee what future discoveries will reveal. Notably, many viewpoints within neuroscience consider the accounts of experiences given in the mental vocabulary of the first-person perspective as valuable research data. It seems in fact that with regard to the study of mind and brain, we have to be committed to a dualist epistemology. Thus, although neuroscience will be able to prove that the mental stuff is not different from the physical one, probably our stubborn habit of speaking about beliefs, hopes, fears, intentions, and so on will hardly be eradicated.

Essential Bibliography

- Churchland PM (2012) *Plato's camera: how the physical brain captures a landscape of abstract universals*. MIT Press, Cambridge, MA.

The book develops the hypothesis that the brain acquires knowledge about the world through a complex process in which high-dimensional vector maps are constructed within a cognitive space. The discussion is at times quite technical; however, this book offers a perfect example of how fruitful philosophy can be when it meets science.

- Churchland PM (2013) *Matter and consciousness*, 3rd edn. MIT Press, Cambridge, MA.

This is a wider and updated edition of Churchland's classic book written in favor of eliminative materialism. With a plain and engaging style, the author discusses the most important philosophical approaches to the mind-body problem by highlighting their pros and cons.

- Churchland PS (2002) *Brain-wise: studies in neurophilosophy*. MIT Press, Cambridge, MA.

With a clear and attractive style, the author introduces the reader to the field of neurophilosophy and examines the classical topics of philosophy of mind within the framework of neuroscience.

- Churchland PS (2011) *Braintrust: what neuroscience tells us about morality*. Princeton University Press, Princeton.

The author discusses the thought-provoking approach of neuroscience to ethics. This expanding and intriguing field of research, which is called neuroethics, aims to account for the origin and nature of morality and ethical values in neurobiological terms.

Intentionality as the Hallmark of the Mental

I shall give reasons for thinking that consciousness is a form of intentionality, the mind "direction upon its objects."
(Aspects of Psychologism)

Tim Crane (born 1962) is an English philosopher who mostly works within the fields of philosophy of mind and metaphysics. He obtained his BA from Durham University, his MA from the University of York, and in 1989 his PhD from the University of Cambridge. From 1990 to 2009, he taught at University College London. He was also director of the Institute of Philosophy in London between 2005 and 2008. Since 2009, he has been appointed as Knightbridge Professor of Philosophy at the University of Cambridge.

Crane's philosophical contributions to the philosophy of mind are widely recognized and support the view that the mind cannot be explained in physical terms only, that experience can have nonconceptual contents, and that the property of being intentional is what distinguishes mental states from physical ones. With regard to the last point, Crane puts this tenet at the heart of his philosophy of mind, which he defines as a particular type of *psychologism*. This term – as it is used within the tradition of the analytical stream of philosophy inaugurated by Frege, Russell, and Wittgenstein, as well as within the phenomenological tradition inaugurated by Husserl – designates the attempt to explain logical, mathematical, and normative entities in psychological terms. The consequences of this view are far-reaching as it discards the human mind as the privileged seat in which thought occurs. In other words, there is a preeminence of the structure of logic, mathematics, thought, and language over the structure and organization of the mind. In this sense, the former does not derive from or depend on the latter, so that thoughts are not to be considered as the products of the mind as well as the life of meanings does not entirely take place within our heads (see also Chap. 13 about the claim that meanings are not in the head).

From the viewpoint of cognitive neuroscience, this position can be seen as bizarre. Neuroscientists are used to explain all kinds of human behavior, including logical and moral reasoning as well as mathematical thinking, in virtue of cerebral mechanisms and processes. According to the mainstream neuroscientific approach, every aspect of human life, from actions to ideas, should find its neural correlates within the brain (about what it is a neural correlate and, in particular, the quest for the neural correlates of consciousness, see Chap. 17). To concede that thoughts are other than the products of brain functioning is to introduce a dualistic assumption in

our scientific picture of the world, which is highly problematic. All those thinkers who charge neuroscientists of psychologism should also explain what, according to them, thoughts, logical, and mathematical entities are made of and how these entities can interact with or at least be grasped by brain processes (see Chap. 6 for a discussion of the classical dualistic position within the philosophy of mind).

Crane accepts the critique of the philosophical tradition but, at the same time, argues in favor of the independence and importance of the mind by claiming that mental entities are self-standing parts of reality, which are susceptible to different kinds of analysis, from the phenomenological to the empirical to the conceptual ones. In this sense, he puts forward a weak version of psychologism. In doing so, he identifies a criterion which helps distinguish what parts of reality are mental and what parts are not. The criterion is summed up in the phrase “intentionality is the mark of the mental,” a thesis that was already held and defended by nineteenth-century German philosopher Franz Brentano.

Intentionality is a technical philosophical concept that derives from a Latin root whose meaning is “tending to.” Mental states are therefore thought to be *intentional* insofar as they are *directed to something*. Intentionality has become a key concept within the philosophy of mind, although the terminology is unfortunate because it seems to be related to another meaning of the word “intentional,” which is “to do things on purpose.” However, the two meanings are radically different and are not to be confused. So, when we say that an action was intentional, we are saying that this action was meant to be done on purpose, whereas when we say that a mental state is intentional, we are saying that this mental state is directed to a certain object or is about something (“tending to something”).

According to Brentano and Crane, mental states present *intentional inexistence*, which is to say that they include something as an object within themselves. In this context, the term “inexistence” does not refer to objects that do not exist (which in certain cases can be true) but rather to the idea that objects *virtually exist* or *are represented in* the mental states themselves. For instance, in hearing a sound such as a car horn, the sound which is heard (i.e., the physical event) is contained within the act of hearing it (i.e., the mental event). Similarly, when we have an idea something is conceived, when we fall in love something is loved, when we develop a desire something is wanted, etc. Arguably, the use of the term “inexistence” in this context may lead to ambiguity, as the same idea could have been conveyed through the expression *intentional existence* or *intentional inexistence*.

It has been argued by other philosophers that intentionality does not characterize all the mental states but only a part of the category. According to this view, there are mental events, such as beliefs, fears, hopes, and desires, which have an intentional nature, because, as we have seen, in each of these cases, something is believed, feared, hoped, and desired. Conversely, other mental events, such as pain and states of nervousness, elation and anxiety, do not present with intentionality, because they do not seem to be about anything in particular. In contrast to this common view, Crane argues that the property of being intentional should be thought of as the hallmark of the whole mental realm.

Based on his analysis of Brentano’s work, Crane claims that also sensations, such as pain and pleasure, are intentional. In fact, the objects of intentional mental

processes can be either external or internal phenomena. When it comes to sensations, the mind can be thought of as directed to something: the sensation itself. We can therefore say that in the process of sensation, something is sensed, as well as in the process of pain, something is felt to be in pain: a body part. Similarly, also states of anxiety and nervousness can be about or directed to something. Crane suggests that being anxious in this way is a matter of thinking of oneself as being in a certain position in the world; it is to regard the world, in other words, as a potentially disturbing, perilous, or frightening place for oneself. In this sense, Crane endorses twentieth-century French philosopher Jean-Paul Sartre's motto that *emotion is a specific manner of apprehending the world*.

Crane identifies two main features of the concept of intentionality. The first feature is a relational structure: being in a mental state always means having a representation *of* something. In the case of consciousness, this aspect is expressed by saying that consciousness is always consciousness *of* something. The second feature is the perspectival nature of intentionality, which is called by American philosopher John Searle (see Chap. 15) *aspectual shape*. The aspectual shape is the particular way the object is presented in the mental state. In other words, it is not possible to think about something but in a certain way, by depicting it with certain properties or characteristics as well as with certain linguistic connotations. For instance, we can think of the moon as the satellite of the Earth, or as a world with no atmosphere, or as the most brilliant celestial object in the night sky, or as a goddess who fell in love with Endymion, etc.

According to Crane, intentionality therefore results in a tripartite relationship: an intentional content (the object) is related to a subject (the mind) by an intentional mode (the aspectual shape). Within this picture, consciousness appears to be a form of intentionality, which can be described as an emergent property of the brain. Based on their intentional structure, mental states as well as conscious experiences literally emerge from a specific physical organization, in a way that cannot be predicted from the properties of their underlying physical underpinnings. Similarly, the waves on the surface of the sea appear to have a life of their own, moving and combining in ways transcending the rules which govern a single molecule of water (Fig. 3.1).



Fig. 3.1 Pieter Bruegel the Elder (circa 1525–1569), *The Storm at Sea* (1568), oil on panel, Kunsthistorisches Museum, Vienna (Image cropped by authors)

Thus, although consciousness depends on a physical substrate, it is nonetheless distinct from physical properties and cannot be explained in physical terms only. This is entirely due to its intentional structure, which according to Crane can also account for the qualitative or phenomenal aspect of consciousness. As we shall see in Chap. 23, the concept of intentionality might play an important role in the neuroscientific study of consciousness, as a further testimony to the value of the dialogue between neuroscience and philosophy of mind.

Essential Bibliography

- Crane T (2001) *Elements of mind: an introduction to the philosophy of mind*. Oxford University Press, New York.

In this book, Crane provides the reader with a lucid introduction to the main issues within the philosophy of mind, including the nature of intentionality or mental representation and the nature of consciousness and perception. He also proposes that the property of intentionality might be the common aspect to all mental processes as well as the basis for the emergent phenomenon of consciousness.

- Crane T (2003) *The mechanical mind: a philosophical introduction to minds, machines and mental representation*. Routledge, London.

This is an insightful and clear book about the mind-body problem and the nature of consciousness from the perspective of a mind regarded as a mechanical device capable to combine different mental representations. This combinatorial cognitive model is thoroughly discussed and pushed to its limits.

- Crane T (2014) *Aspect of psychologism*. Harvard University Press, Cambridge, MA.

This book is a collection of essays exploring fundamental philosophical questions regarding consciousness, perception, and intentionality. Crane illustrates his version of psychologism, which considers intentionality as the hallmark of all mental phenomena.

A Mismatched Couple of Vocabularies for the Same Reality

*Even if someone knew the entire physical history of the world,
and every mental event were identical with a physical,
it would not follow that he could predict or explain
a single mental event (so described, of course).*

(Essays on Actions and Events)

Donald Davidson was one of the greatest analytic philosophers of the second half of the twentieth century. His ideas have been influential and stirred important debates across philosophy of language, epistemology, and philosophy of mind.

Davidson was born on 6 March 1917 in Springfield, Massachusetts, and died on 30 August 2003 in Berkeley, California. His life was intense and full of interests. He had a broad education, ranging from literature to arts, architecture, and music. He could ride a horse, pilot a plane, and play piano (he used to play four-hand piano with Leonard Bernstein when they were student at Harvard). He liked surfing, climbing, and traveling. During World War II, he joined the navy and was trained to teach gunners how to recognize the moving shape of enemy planes. In 1946, he came back to Harvard and dedicated himself to the study and teaching of philosophy in prestigious American universities, like Stanford, Princeton, and Berkeley.

Davidson's philosophical essays are woven with rigorous argumentation. His writing style is characterized by logical rigor and direct focus on the possible solutions to the problems which are under discussion. The intrinsic originality of his thought often leads him to interesting counterintuitive positions, including the influential ideas originally published in 1970 in his famous essay on *Mental Events*. In this paper, Davidson endorses an ingenious theory about the relationship between the mental and the physical, called *anomalous monism*. As suggested by the theory's name, Davidson argues that although reality is made by only one ingredient (i.e., matter) and is therefore subject to the natural laws, it seems however to behave *anomalously*. Actually, even if we knew the whole physical history of the world – Davidson claims – we could not be able to predict or explain our mental life.

This is possible because anomalous monism introduces a disconnection between ontology, the way the world is really made, and epistemology, the way we know how the world is made. From the ontological point of view, nature is intimately physical. However, from the epistemological point of view, we cannot help but describe reality as composed of two different kinds of events, that is, mental ones and physical ones, by using two different sets of vocabularies that are inevitably

mismatched. This sort of mismatch implies that sentences expressed in mental terms and sentences expressed in physical terms are not equivalent, since there are no natural laws or nomological rules (hence the *a-nomalousness* of monism) which can be applied in order to reduce the former to the latter.

More analytically, Davidson's argument is based on the acceptance of three epistemological principles: the interaction principle, the cause-law principle, and the anomalism principle.

The first principle states that *at least some mental events causally interact with some physical events*. It seems in fact very plausible to assume that occurrences of certain mental events take place whenever other occurrences of certain physical events take place, and vice versa. For instance, the desire to eat chocolates is accompanied by both some physical states in the body and the fact that there is a box of chocolates in the kitchen. Similarly, the fact that a box of chocolates was bought and stored in the kitchen is related with the belief that chocolate is tasty and has a good influence on one's mood.

The second principle, asserting the nomological character of causality, states that *events related as cause and effect fall under strict laws*. In this context, "strict" means that the laws are to be expressed in fully articulated antecedents, which logically lead to fully articulated consequences. Similarly to the first principle, this assumption is highly intuitive and permits to construct causal chains of events which can consistently explain natural phenomena; in addition, if coupled with the interaction principle, it seems to lead to the conclusion that there can be psychophysical laws which strictly govern the relationship between mental and physical events. In fact, if a particular mental event m_1 is related to a particular physical event p_1 , then, given the two principles stated above, there should be a strict law such as $P_1 \rightarrow M_1$, so that, whenever events of kind P_1 occur, events of kind M_1 must follow.

Davidson, however, denies the existence of such psychophysical laws in his third principle, which explicitly states the anomalism of the mental: *there are no strict laws on the basis of which mental events can predict, explain, or be predicted and explained by other events*. Therefore, Davidson's mental anomalism refutes the possibility of any strict laws in which mental predicates can figure, including laws formulated with mental predicates only, such as $(M_1 \text{ and } M_2) \rightarrow M_3$, as well as laws formulated with mental predicates in either the antecedent or consequent, such as $(M_1 \text{ and } M_2) \rightarrow P_1$ or $(P_1 \text{ \& } P_2) \rightarrow M_1$, or mixed versions of these forms, such as $(P_1 \text{ \& } M_1) \rightarrow M_2$. In particular, Davidson rejects the enunciation of any set of laws expressed in the form $P_1 \leftrightarrow M_1$.

Davidson's anomalous monism refuses any kind of strict nomological relationship between mental and physical events, in agreement with models of "token identity theory" between mind and brain. In principle, identity between mental and physical events might be of two kinds, one of particulars (token identity), the other of categories (type identity). Let us consider the word "apple." In this word, there are five tokens of letters, but only four types of letters, because the "p" is repeated twice. Similarly, with regard to mental and physical events we can refer to either tokens or types. Type identity theory maintains that all the mental events of a certain type are identical with all the physical events of a certain type: every occurrence of

“p” in “apple” or in any other word containing the letter “p” will be always the same. Token identity theory, in contrast, maintains that every occurrence of a particular mental event is identical with a specific occurrence of a particular physical event. In this case, both mental and physical events are unique and unrepeatable items of the world’s furniture: occurrences of “p” in any word containing this letter will never be exactly the same, such as in “apple.” Although philosophers have put forward interesting arguments in favor of either type or token identity theories, it is possible that the problem of the exact relationship between mental events and brain states will have to be empirically addressed by neuroscience.

In characterizing events as particulars, Davidson (1985) accepts the definition of “event” proposed by Quine, according to whom an “event” is everything taking place in a framework of spatiotemporal coordinates. According to this definition, two events are identical if and only if they occupy exactly the same location in space and time. This conception of identity provides solid grounds to Davidson’s assertions that “there cannot be two events alike in all physical respects but differing in some mental respects” (Davidson 2001a, 214) and that “if two events fail to share a mental property, they will fail to share at least one physical property” (Davidson 1995, 266). Thus, any changes in a person’s mind must be accompanied by corresponding physical changes in his/her brain.

Davidson’s anomalous monism is therefore open to the possibility that mental properties supervene upon physical ones, that is to say that mental events might be entailed by or consequent to the existence or establishment of physical events. Davidson’s endorsement of supervenience emphasizes a feature of anomalous monism which has been repeatedly criticized: the explanatory primacy of the physical over the mental. This implies that mental events have an *epiphenomenal* nature, that is, they are the mere shadows of the corresponding physical events occurring in the brain. Davidson’s (1993) reply to the objection that anomalous monism inevitably leads to the epiphenomenalism of mental properties is that *causal relations* should not be confused with *causal explanations*. Causal relations exist among events independently of how we describe them. Causal explanations, in contrast, are confined to an epistemological framework where rational descriptions of human behavior correlate only with other rational descriptions of human behavior. Our way of speaking about mental events, states, and properties, by using a vocabulary that refers to intentions, beliefs, and desires, does not compel us to accept ontological consequences in favor of the real causal power of the mental, even though it provides us with the only conceptual apparatus suitable for giving rational accounts of human action. It is not clear whether Davidson’s reply may stand the charge of epiphenomenalism. However, it seems true that, from an epistemological perspective, we are compelled to describe different aspects of the same reality using two different vocabularies. Importantly, these descriptions or rational explanations, Davidson warns, should not lead us to think of them as referring to authentic causal relations.

Davidson’s anomalous monism is based on complex ontological and epistemological considerations, oscillating between a metaphysical doctrine and a theory of knowledge. In fact, in Davidson’s philosophy, questions about the ontological

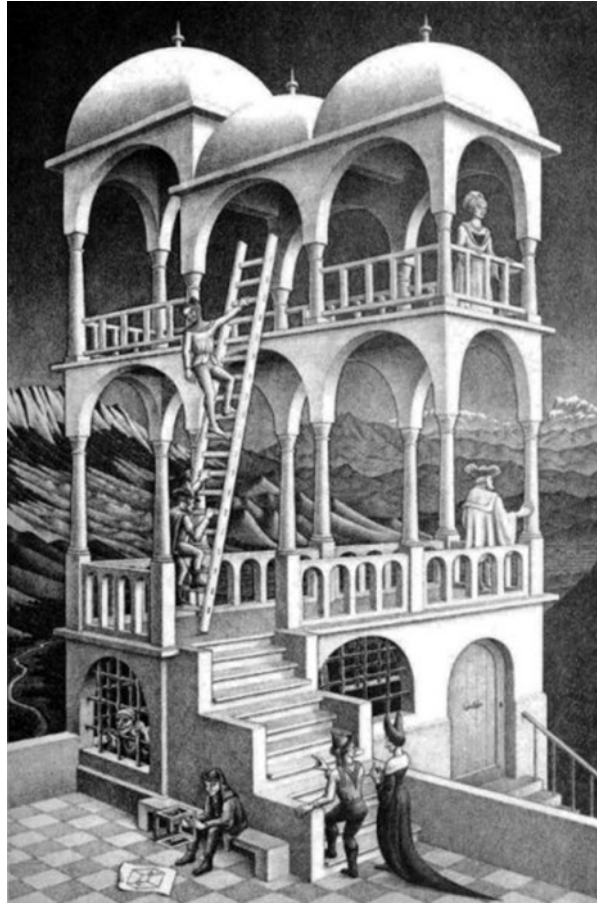
status of events are always mixed with questions about the logical structure of sentences which refer to these events. In particular, Davidson takes into consideration sentences on mental events which are characterized by intentionality – the property of being directed to something else – such as beliefs, desires, intentions, hopes, etc. (but, in principle, the same considerations can be applied to conscious mental events). The most fundamental aspect of these expressions is that it is not possible to have one without having a host of others. For instance, one's desire to eat chocolate is connected to the belief that there is a box of chocolate in the kitchen, which is, in turn, linked to the belief that the kitchen is downstairs, etc. As a consequence, mental predicates cannot exist independently of the whole or cannot be understood without reference to the whole, a thesis known as *holism*. What is more, this seemingly infinite interconnectedness among mental properties has led to the idea, endorsed by Davidson, that the mental domain must be open and unpredictable, in contrast to the physical domain, which should be closed and deterministic.

According to this line of thought, anomalous monism would constitute an original attempt to save free will and the autonomy of human action from the constraints of determinism, as it aims at reconciling the mechanistic view of physics with our natural tendency to speak about beliefs, intentions, and desires as causes for our behavior. However, anomalous monism fails to give any conclusive reason as to why we should take as facts both that there is no difference between mental and physical events *and* that the former cannot be in any way reduced to the latter. It is therefore unclear how anomalous monism could provide a sound theoretical framework for a fruitful dialogue with neuroscientific research programs.

In a sense, Davidson's theory could be exemplified by a work of M. C. Escher (Fig. 4.1).

In the lithograph named *Belvedere*, we see a turret composed of three floors. At first glance, nothing seems unusual about the building. The first floor is a solid and closed compartment with barred windows, while the middle and the top floors are open at the sides with arches and pillars supporting a roof with three little domes. The turret, however, is an impossible structure in a three-dimensional space. In fact, the pillars of the middle floor crisscross each other, so that they appear to join the front of the top floor to the back of the middle floor. In addition, Escher places a ladder that connects the inside of the middle floor to the outside of the top floor. Thus, the composition eventually produces a bewildering effect. Anomalous monism might be illustrated by the following analogy, which is based on a free interpretation of Escher's work. The first floor may represent the stable and closed physical realm and the top floor the open and unpredictable mental realm. The middle floor may symbolize the impossibility of a reduction of mental events to physical events by means of strict natural laws (the pillars). Still, the ladder can be considered as the rational explanations which *apparently* guide to the top floor of the airy mind.

Fig. 4.1 M. C. Escher (1898–1972), *Belvedere* (1958). (M.C. Escher’s “Belvedere” © 2014 The M.C. Escher Company-The Netherlands. All rights reserved. www.mcescher.com)



Essential Bibliography

- Davidson D (2001a) *Essays on actions and events*, 2nd edn. Clarendon Press, Oxford.
The book collects some seminal essays (including *Mental Events*) on the nature of human action that have become classics in the fields of philosophy of language, mind, and psychology. A wide variety of topics are discussed, ranging from the freedom to act to the weakness of the will, the logical form of talk about actions, intentions, and causality.
- Davidson D (2001b) *Subjective, intersubjective, objective*. Clarendon Press, Oxford.

The volume is a collection of epistemological essays which explore the relationship between knowledge, mind, and language. The insightful discussion begins with the analysis of how one can have knowledge of his/her own mental states, moving to the knowledge of other minds, and finally to the knowledge of the external world.

- Davidson D (2004) *Problems of rationality*. Clarendon Press, Oxford.

This posthumously published book focuses on the complex theme of rationality, investigating the conditions for the attribution of mental states to other people and species, our understanding of value judgments, and the problems presented by seemingly irrational thoughts and actions. The volume is enriched by an introduction by Marcia Cavell, Davidson's widow and daughter of philosopher Stanley Cavell, and by a fascinating interview with Davidson himself.

The Multiple Drafts Model

...consciousness [...] is not a special “medium of representation” in the brain into which content-bearing events must be “transduced” in order to become conscious. It is rather a matter of content-bearing events in the brain achieving something a bit like fame in competition with other fame-seeking [...] events.
(Sweet Dreams)

American philosopher and cognitive scientist Daniel Dennett (born on 28 March 1928 in Boston, Massachusetts) is one of the most important theoretical thinkers in the field of philosophy of mind and consciousness studies. Dennett’s philosophical contributions are far-reaching, as they include the problem of free will and evolutionary theory. He graduated in philosophy from Harvard, where he was a student of philosopher and logician W. V. Quine, mentor of Donald Davidson (see Chap. 4), and received his PhD from the University of Oxford, where he studied with Gilbert Ryle. Ryle was editor from 1947 to 1971 of the philosophical journal *Mind* and achieved international fame for his book *The Concept of Mind*, published in 1949, in which he criticized Descartes’ theory (see Chap. 6) which considered mind and body as two separate entities. Interestingly, Ryle referred to this distinction as a “categorical mistake.” Daniel Dennett is currently Co-director of the Center for Cognitive Studies and Austin B. Fletcher Professor of Philosophy at Tufts University.

Within the philosophy of mind arena, Dennett is famous for arguing that qualia – the qualitative aspect of a conscious experience – do not exist. He analyzes what the properties of *qualia* are supposed to be and finds that these kinds of properties or features are vague and ill defined. Dennett examines the most important among these properties: intrinsicality. A *quale* (singular of *qualia*) is thought to be *intrinsic* to a mental state. But if so, there should be something essential and invariable in the nature of a specific *quale* that does not depend on external elements. However, philosophers commonly think that the same *quale* (say, seeing a pink tie) can be different for every human being. For instance, there are substances (such as phenol-thio-urea) that have a bitter taste to a high proportion of people, but no taste at all to others. Moreover, the same *quale* can change through time even for the same person. Adults can love the same food they disliked when they were children (acquired taste). Therefore, Dennett concludes that the supposed intrinsic nature of *qualia* is, on the

contrary, extrinsic and relational. Following this line of reasoning, he resolutely denies the existence or importance of *qualia* as ineffable and completely private subjective experiences. Thus, the so-called *hard problem* of consciousness evaporates, along with the hypothetical conceivability of zombies (see Chap. 1).

From this theoretical approach, Dennett draws another important conclusion. If our subjective experiences are not strictly private, but partly extrinsic and relational, then we can collect descriptions of “what it is like” to have them under various controlled conditions. This method, which is called *heterophenomenology* by Dennett (“hetero-” is a term coming from Greek whose meaning is “other”; Dennett uses it to indicate that the phenomenological inventory is not done by the person who is actually having the subjective experiences but by a trained scientist), provides a catalogue of what people believe to be true about their conscious experiences. This heterophenomenological catalogue, together with the experimental data neuroscientists can gather regarding their participants’ brains and the surrounding environment, encompasses the total set of elements a theory of human consciousness should be able to explain.

The interpretation of these data requires adopting what Dennett calls the *intentional stance*. The intentional stance is the working hypothesis that people are agents whose actions are guided by rational beliefs and desires. For instance, the intentional stance warns the experimenter to take precautions in order to prevent subjects from having experiences that might lead them to irrational beliefs or desires potentially biasing their responses or distort the experimenter’s interpretation of their actions. Consequently, subjects are kept in the dark about the real scope of the experiment, but at the same time, they are given information to correctly understand their task. Dennett claims that with heterophenomenology, he is not suggesting a new method, but simply describing and defending the standard methods already adopted by neuroscience. These methods, if correctly understood and followed, should not leave any conscious phenomena outside the scientific inquiry.

According to Dennett, there is nothing really special or mysterious in having a first-person perspective, because subjective experiences can be accounted for by using the method of heterophenomenology within the framework of neuroscience. To think otherwise would be as believing that inside the brain, there could be a privileged place where everything comes together in a conscious experience. In this privileged place, an “I” or a sort of homunculus would watch all the mental representations like a spectator in a theater (Fig. 5.1).

Dennett calls this view “Cartesian materialism” and the privileged place in the brain “Cartesian theater.” Accordingly, a conscious experience would occur when the homunculus witnesses the representation of the content of that experience. However, Dennett strongly argues that this kind of view leads to a mistaken and incoherent theory of conscious experience. In reality, in the brain, there is neither a Cartesian theater nor a homunculus capable of becoming conscious of mental representations.

In fact, from the functional point of view, the work that the homunculus would do in the Cartesian theater can be subdivided and allocated in time and space to specialized brain modular agencies. Once the information has been elaborated by one of these specialized agencies, this elaboration does not have to be performed again in a central representational system, nor does it have to be transferred somewhere and represented again in order to enter memory. Moreover, the brain presents

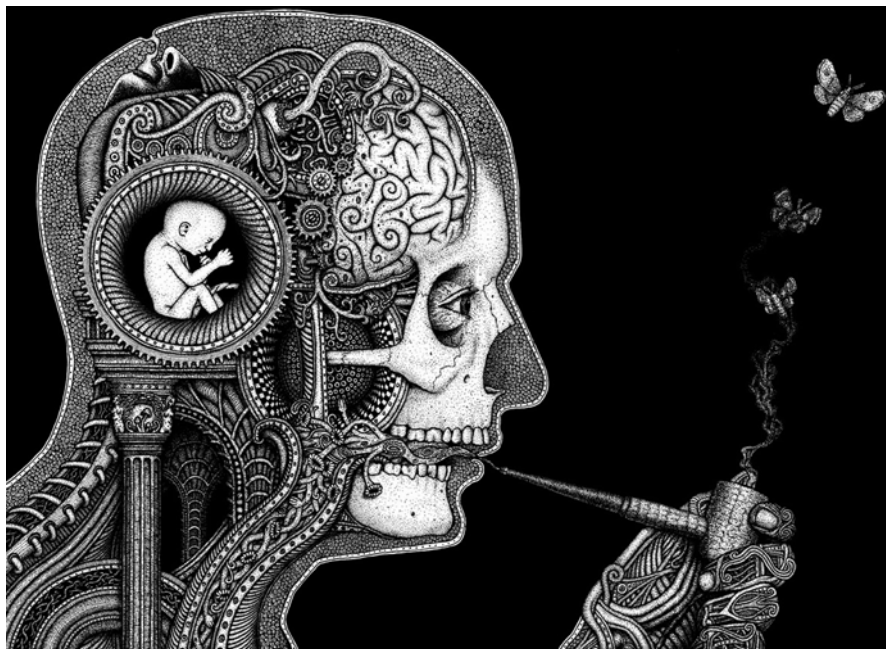


Fig. 5.1 Anonymous. Fantastic representation of the homunculus inside the head

a massive parallel architecture, in which multiple (and often rival) processes and streams of informational contents take place both simultaneously and asynchronously. Dennett calls these processes *multiple drafts* and the theory of consciousness which derives from this approach *multiple drafts model*.

In Dennett's view, at any time multiple competitive drafts try to become predominant in the brain until a "final draft" appears. In his words, a draft becomes the final one when it gains more "fame" in the brain compared to others. To achieve fame means that the final draft will last longer and have more sequelae (i.e., influences over other brain processes) than the others, entering a *global workspace* (see Chap. 16 for the concept of "global workspace") and thereby becoming largely available. This global availability does not in turn cause a further effect that would ignite consciousness; on the contrary, it is itself a conscious state. But just as becoming famous within society is not a precise datable event, so the exact moment in which a draft achieves fame into the brain is not a precisely detectable transition. This is the reason why Dennett claims that the question of when we become aware of something is ill defined. To better explain this important point, Dennett invites us to think about the phenomenon of coming to notice that a clock is chiming. Sometimes it happens that only on the third or the fourth peal we realize that the clock is chiming; however, we can retrospectively count the chimes back in time with the help of our conscious memory. But when exactly were we first conscious of the first chime? Did we become conscious of that chime when it actually occurred or when we recollected it in our memory? According to Dennett, these are questions with no precise answers.

Thus, Dennett argues that consciousness is neither intrinsic nor dispositional. It is, instead, a phenomenon of actualization of specific informational contents whose drafts are able to win the battle for “cerebral celebrity.” This leads to a simple demarcation between conscious and unconscious processes: the former are those that are widespread and “famous” throughout the brain; the latter are those that remain confined to a specialized brain agency and do not gain enough fame.

Dennett holds a strictly functional approach to consciousness. In his view, consciousness and every other mental function do not essentially depend on their neural substrate but exclusively on brain activity. Thus, if a machine could replicate the functional organization of the human brain, it would also be able to achieve full-blown consciousness. Artificial consciousness seems in principle possible, but only future empirical research will tell us whether or not it is also feasible. However, neuroscience shows that the relationship between brain anatomy and brain function is not to be underestimated: neuroanatomy matters because it allows specific functions, which in turn can be achieved because of certain neuroanatomical connections. It is therefore more likely that we will be able to build a conscious machine when we achieve a deeper understanding of how the brain works rather than the other way round. What is more, it seems not possible to thoroughly understand mental functions without understanding their meaning, which appears to be deeply rooted in neuroanatomical connections. It can be argued that consciousness might have to rely on a structure which is modeled on ours. To use a metaphor, both airplanes and birds can fly, but in order to fly like a bird, you must possess the characteristic structure of its wings.

Essential Bibliography

- Dennett DC (1987) *The intentional stance*. MIT Press, Cambridge, MA.
In this book, Dennett proposes and develops the idea of the *intentional stance*, the view according to which the behavior of human beings is to be considered as rational and based on mental states, such as intentions and beliefs. Dennett argues that this is the attitude we implicitly have every time we interact with others; it also provides us with a lens through which we can see and interpret human actions. Without this lens, there would be no mutual understanding.
- Dennett DC (1991) *Consciousness explained*. Little, Brown, Boston, MA.
This book presents and discusses Dennett’s multiple drafts model of consciousness. Dennett criticizes the view that inside our brain, there is an inner “I” or a homunculus which brings to consciousness what it is represented in the mind. The problem of *qualia* and the method of heterophenomenology are thoroughly analyzed, leading other scholars to claim that Dennett “explained consciousness away.”
- Dennett DC (2005) *Sweet dreams: philosophical obstacles to a science of consciousness*. MIT Press, Cambridge, MA.

Almost all the themes presented in *Consciousness Explained* are here revisited and discussed in the light of the philosophical debates that raged along the years between the two books.

Mind and Body: Two Different Substances

*If I had merely ceased thinking
[...] I should have had no reason to believe that I existed.
From this I knew I was a substance whose whole essence or nature
is simply to think, and which does not require any place,
or depend on any material thing, in order to exist.
(Discourse on the Method)*

René Descartes (Fig. 6.1) is commonly considered the father of the modern philosophy. Undoubtedly, he was one of the greatest philosophers ever lived, as his view of the natural world in pure mechanical terms heavily influenced the course of the Western thought. Specifically, he contributed to the decline of Aristotelian physics and to the rise of modern science. Moreover, as a mathematician he invented the Cartesian (from his Latin name *Cartesius*) coordinate system that made possible the development of analytic geometry.

Descartes was born on 31 March 1596 in France and died of pneumonia on 11 February 1650 in Sweden. He studied logic, morals, physics, metaphysics, and mathematics at the Jesuit College of La Fleche. His family wanted him to be a lawyer, so he moved to Poitiers to study law and obtained the degree in 1616. However, in his life, he never practiced as a lawyer: he became instead a gentleman soldier in the army of Prince Maurice of Nassau.

During the night of 10 November 1619, while the army stationed in Neuburg, Descartes had three dreams that – in his own words – changed the course of his life. In the first, he was in a street in the company of some friends. One of his legs was unsteady, so that, unlike his friends, he had to struggle to remain standing against the blowing wind. In the second dream, he had the impression to be awakened by a violent thunder and saw his room bursting with sparks. In the third, an old man gave him a book of poems in which he read the words: “What way of life shall I follow?”

Descartes took the first dream to mean that he should have to be skeptical of all the learning he had received in his education. The second dream was interpreted as a metaphor of the glaring light of truth. The third vision convinced him that a way for truly understanding the nature of reality was possible. Accordingly, he decided to dedicate himself to the reform of the human knowledge.

Descartes set the problem of the mind-body relation as we are still acquainted with. His position, known as Cartesian or classical dualism, holds that the human

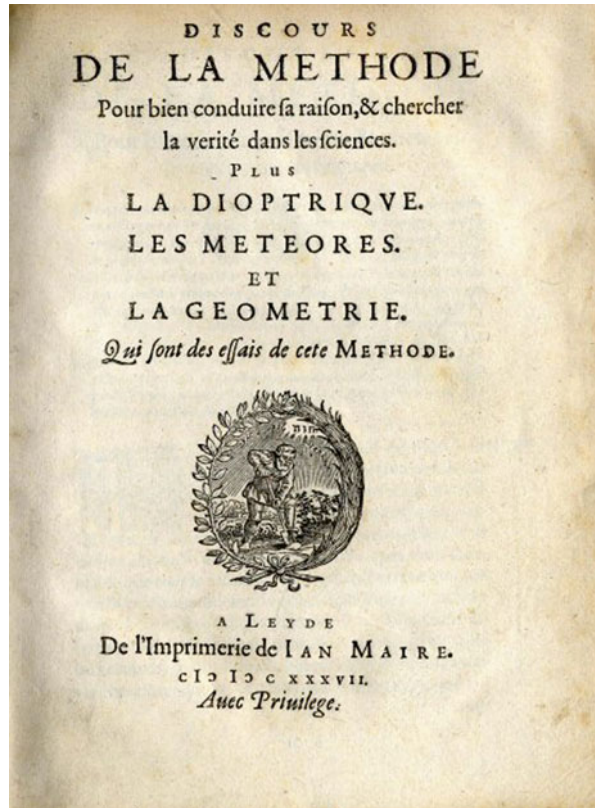
Fig. 6.1 Frans Hals (1580–1666), *Portrait of René Descartes*, oil on canvas (circa 1649), Musée du Louvre, Paris, France (Image cropped by authors)



beings are composed of two substances, one which is purely material or corporeal, the *res extensa*, the other which is purely immaterial or intellectual, the *res cogitans*. Human beings are the only creatures in the world to be born with a soul. Animals, on the contrary, are made by one ingredient, that is, matter, and thereby can be exclusively described in mechanical terms. The claim that the universe as a whole, from the celestial bodies to animals (except humans, of course) is a huge clockwork that perfectly obeys to natural laws is one of the ideas that shaped modern science. With a clear-cut dichotomy between two realms of reality, Descartes made it possible to systematically investigate nature by means of mathematics.

It was a great achievement of the human intellect to believe that the material universe could be explained in a mathematical and mechanistic way. However, this point of view posed two serious problems. On the one hand, human behavior and functions would have remained unaccountable by science; on the other hand, there was the need to find out how and where in the human body the two substances (mind, or *res cogitans*, and matter, or *res extensa*) could interact. These questions – which are of fundamental importance both for the neuroscience and for the philosophy of mind and psychology – were addressed by Descartes in several works. In particular, three books stand out: the *Discourse on the Method*, published in French

Fig. 6.2 Title page of *Discourse on the Method*, first edition (1637)

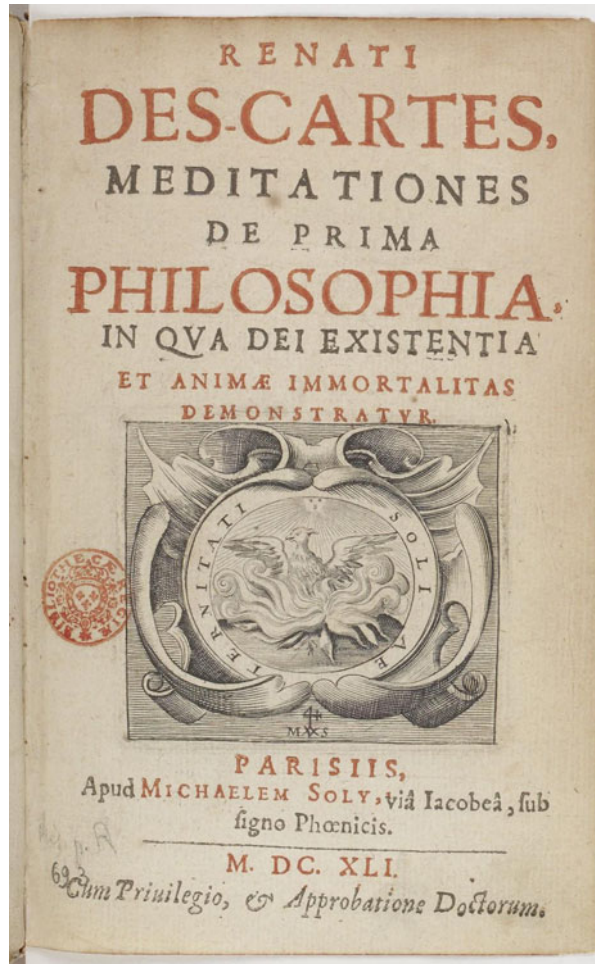


in 1637; the *Meditations on First Philosophy*, published in Latin in 1641; and the *Passions of the Soul*, published in French in 1649.

In the *Discourse on the Method* (Fig. 6.2), Descartes declares to never accept anything for true unless he does clearly know to be such. This is the renowned method of doubting everything until one is completely certain, with the help of a clear and distinct acquaintance that the thing he or she is conceiving of is absolutely real. Applying the method to himself, Descartes is eventually able to find something that is not susceptible to doubt. He suddenly realizes, in fact, that he cannot put in doubt himself while he is doubting. As a result, he concludes that a thinking being must be real. This conclusion is expressed by the famous remark “I think, therefore I am.” Furthermore, based on the conceivability of a substance whose nature is simply to think (without the need of any corporeal attribute in order to exist), Descartes is inclined to take for granted that the mind and the body are different substances.

The question regarding the nature of mind is tackled more analytically in the *Meditations on First Philosophy* (Fig. 6.3). Specifically, the difference between mind and matter is addressed in the Second and in the Sixth Meditations. In these Meditations, Descartes asks himself what kind of thing is a thinking being. The answer is that a thing that thinks is a thing that doubts, understands, wills, affirms,

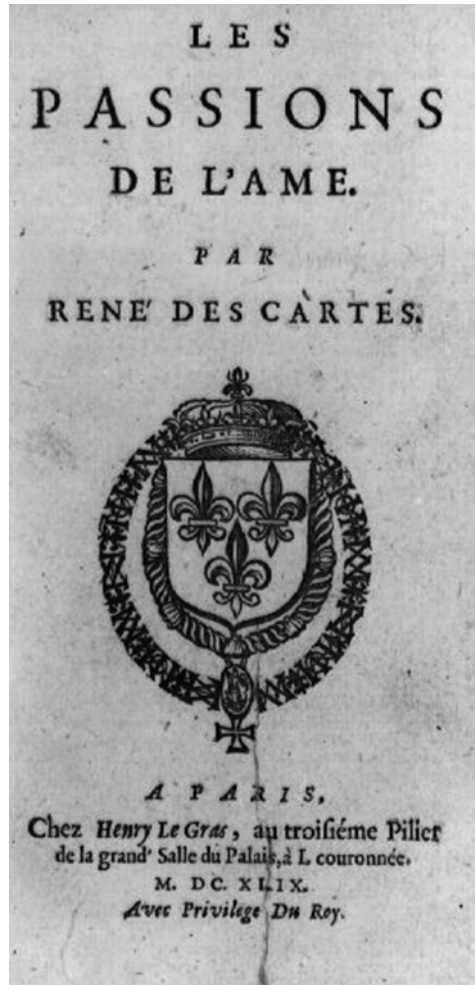
Fig. 6.3 Title page of *Meditations on First Philosophy*, first edition (1641)



denies, refuses, has mental images, and eventually seems to have sense perceptions. A thinking being is, in a nutshell, an intellectual substance whose very essence is to represent by means of ideas the outside world. This theoretical position disconnects the mind from the world and leads Descartes to the truth, which is, according to him, self-evident, that mind and matter are to be diverse and, what is more, separate substances in so far as they can exist independently of one another. The two substances are thus characterized by opposite attributes: matter is extended throughout space, while mind is not, being wholly unextended.

In Descartes' philosophy of mind, consciousness appears to be the hallmark of the mental. He actually holds that mind is the substance in which thought resides (*Meditations on First Philosophy*, 7:161) and affirms with regard to the term "thought" that it applies to "everything that is within us in such a way that we are immediately aware of it" (*Meditations on First Philosophy*, 7:160). In another place,

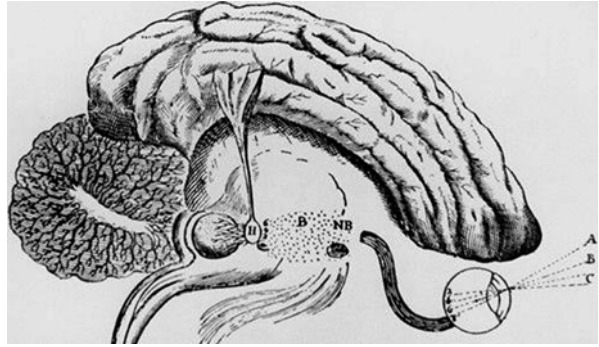
Fig. 6.4 Title page of *Passions of the Soul*, first edition (1649)



he maintains that every thought is somehow conscious (*Meditations on First Philosophy*, 7:226). This aspect of Descartes' metaphysics appears now to be one of the most outdated, since neuroscience and psychology have amassed a bounty of evidence on how a great deal of our mental life unfolds unconsciously. Another aspect that is at odds with the current brain science is of course Descartes' solution to the problem of how two different and separate substances could interact in order to make it possible for the soul to govern the body.

The issue of the interaction between the *res cogitans* and the *res extensa* received an extensive account in the *Passions of the Soul*, the last of Descartes' books published during his lifetime (Fig. 6.4). The book is divided in three parts. The first one deals with passions in general and the nature of man, whereas the other two sections treat specific passions, along with their number and attributes. The Cartesian attempt

Fig. 6.5 Illustration from Descartes' *De homine* (1662) showing the location of the pineal gland (identified by letter H) and its role in the process of vision



to solve the mind-body interaction problem is contained in the first section. Here Descartes explains that the mind (or soul) is jointed to the body as a whole, in virtue of having no extension nor dimensions. Nonetheless, he identifies a point in the body, specifically in the center of the brain, where soul exercises its function more effectively than in the other parts. In this point resides a small organ, the pineal gland, which seems to Descartes the only fraction of the human brain that is not paired. According to the French philosopher, this aspect was of great importance, because it let the twofold impressions of the senses, especially with regard to the images coming from the two eyes, to reach the soul in a unitary form (Fig. 6.5).

Descartes' hypothesis of the pineal gland as the point of privileged contact between mind and body is ingenious but unfortunately fails to give a satisfactory solution of the mind-body problem within a dualistic frame. We now know, in fact, that the pineal gland is involved in the regulation of the circadian rhythm by secreting melatonin and has, like the other structures of the brain, a left and a right side that are mirror images of each other. The truth is that Descartes was never able to give a plausible account of how an incorporeal or unextended substance and a corporeal or extended substance might interact. This very issue was famously raised by Princess Elizabeth of Bohemia, who asked the French philosopher in a letter to tell her how a man's soul, being only a thinking substance, could determine animal spirits so as to cause voluntary actions. Eventually, Descartes admitted to Elizabeth that he did not have a definitive answer (*Correspondence*, 3:694).

Descartes should be credited for producing the first methodologically rigorous, albeit unsuccessful, effort to elucidate in what way the volition of the soul could move the body (the problem – in contemporary philosophy of mind terms – of “mental causation”). In modern physics, the problem is analogous to that of explaining the causal influence of a nonphysical entity upon physical things, without violating thereby the principle of the conservation of energy.

Descartes can be considered a pioneer in the philosophy of mind. In fact, his mental experiment aimed at conceiving a disembodied mind has encountered unprecedented fortune in the following philosophical debate. For instance, Saul Kripke's (1980) argument directed to conceive a possible world in which human beings can feel pain in the absence of C-fibers stimulation is not, in principle, dissimilar from Descartes' conceivability of human beings who are able to think

without brain activity. Further, Chaps. 1 and 13 show how other authors have used the logical frame behind this line of reasoning to develop arguments against both physicalism and identity theories of mind.

Descartes' classical dualism is currently a minority position within the circles of philosophers and neuroscientists, although it continues to receive endorsement by some scholars. A few examples are Hart (1988), Foster (1991), and Swinburne (1997). Other authors – for example, Hasker (1999) and Lowe (2006) – uphold an original view, inspired by Descartes' philosophy, which maintains the separation between mind and body without claiming at the same time that one substance be physical and the other nonphysical. Finally, Chaps. 19 and 26 will show how Cartesian dualism has recently been discussed with deep interest by two leading neuroscientists. Thus, Descartes' dualism has stirred up an immortal legacy across both contemporary philosophy and neuroscience, suggesting that this theory is a never-ending source of thought-provoking debates.

Essential Bibliography

- Descartes R (1637) *Discours de la méthode pour bien conduire sa raison & chercher la vérité dans les sciences. Plus la dioptrique. Les météores. Et la géométrie. Qui sont des essais de cette méthode.* In: Adam C, Tannery P (eds) *Œuvres de Descartes*, 11 vol, edn. Vrin/CNRS, Paris; English translation: *Discourse on the method.* In: Cottingham J, Stoothoff R, Murdoch D, Kenny A (1984–1991) *Philosophical writings of Descartes*, 3 vol. Cambridge University Press, Cambridge (In this book we have used this edition for quotations, which reports the Adam and Tannery pagination in the margins.)
One of the greatest classic of modern philosophy. The French thinker describes in a clear and charming prose the method of systematic doubt and formulates for the first time his famous statement “I think, therefore I am.”
- Descartes R (1641) *Meditationes de prima philosophia.* In: Adam C, Tannery P (eds) *Œuvres de Descartes*, vol 11. Vrin/CNRS, Paris. English translation: *Meditations on first philosophy.* In Cottingham J, Stoothoff R, Murdoch D, and Kenny A (1984–1991) *Philosophical writings of Descartes*, 3 vol. Cambridge University Press, Cambridge.
This treatise provides a more thorough and analytical inquiry into the subject of the previous book, that is, on how can it be possible for man to gain true and well-founded knowledge of himself and of the world. Descartes tries to demonstrate the existence of both God and the individual soul by painstaking philosophical reflections taking place along 6 days. Each meditation develops in 1 day and discusses a specific topic.
- Descartes R (1649) *Les passions de l'âme.* In: Adam C, Tannery P (eds) *Œuvres de Descartes*, vol 11. Vrin/CNRS, Paris. English translation: *Passions of the soul.* In: Cottingham J, Stoothoff R, Murdoch D, and Kenny A (1984–1991) *Philosophical writings of Descartes*, 3 vol. Cambridge University Press, Cambridge.

Descartes offers his most vigorous attempt to clarify the relationship between mind and body. After having investigated the nature of passions and the functions of the soul, he puts forward the pineal gland hypothesis as to how mind and matter can interact. To boot, Descartes presents brilliant observations on various kinds of passions that are of unique psychological interest to this day.

The Language of Thought

*I assume that psychological laws are typically implemented
by computational processes.*
(The Elm and the Expert: Mentalese and its Semantics)

Jerry Alan Fodor (born in 1935 in New York City) is an American philosopher and cognitive scientist. His theories on the nature of language and mental states have heavily influenced both the fields of philosophy of mind and cognitive science. In 1956, Fodor gained his A.B. degree from Columbia University and in 1960 his PhD in philosophy from Princeton University, under the supervision of Hilary Putnam (see Chap. 13). From 1959, he joined the Massachusetts Institute of Technology (MIT), where he taught in the Departments of Philosophy and Psychology until 1986. He subsequently became Distinguished Professor at the City University of New York (CUNY) and in 1988 was appointed State of New Jersey Professor of Philosophy and Cognitive Science at Rutgers University.

Along with philosopher Hilary Putnam and linguist Noam Chomsky, Fodor has strongly criticized behaviorism and significantly contributed to the development of cognitive science. Fodor has theorized and defended a realist conception of those mental states (i.e., beliefs, desires, wishes, hopes, fears, worries, intuitions, etc.) that can be manifested through certain expressions which are called *propositional attitudes* in philosophical jargon. The term “propositional attitudes” derives from the fact that we express the content of these mental states in terms of having an attitude toward a certain proposition. For instance, the sentence “Mary hopes to win the national lottery” implies that Mary holds the attitude of hoping toward the proposition “I will win the national lottery.” These kinds of intentional states form the bulk of what philosophers call “folk psychology,” that is, a collection of common-sense laws linking the contents of one’s mind to one’s actions. Folk psychology is generally used to explain and predict human behavior. Thus, Mary’s hope to win the national lottery will make her buy one or more national lottery tickets and follow the lottery draw on TV.

Fodor has repeatedly argued in favor of both the existence of propositional attitudes and the predictive power of folk psychology. He highlights that every day we all are committed to folk psychological explanations of human behavior, whose

reliability seems to be beyond rational dispute. Everyone who hopes to have chances in winning the national lottery will buy at least a lottery ticket, as well as everyone who is afraid of mice will flee in presence of a mouse.

Within a physicalist framework, propositional attitudes can be thought of under three aspects. They could be accounted for in behavioral terms, if considered as referring to patterns of behavior or to dispositions to behave in particular ways. They could be accounted for in materialistic terms, if considered as identical to types of brain states. Finally, as Fodor suggests, they could be accounted for in functional terms, if considered as relations between mental representations and subjects. Thus, according to Fodor's view, propositional attitudes are to be defined both by what they can cause and by what they can be caused by, that is, by other attitudes, stimuli, and behaviors. In order to achieve such an account, Fodor endorses the idea that human thinking is a kind of computation, which is characterized by a set of rules or instructions that he calls *the language of thought*.

The language of thought hypothesis claims that mental representations have a sort of language-like structure, which is to say that they are made of more basic components (i.e., atomic representations, which cannot be further subdivided in meaningful elements). Just as sentences of our ordinary language are composed of single words, so complex mental representations are composed of simpler atomic representations. And similarly to the meaning of a sentence, which originates from the combination of the meanings of words and their syntactical relations, the meanings of a mental representation are the function of the relationships between its atomic parts and their meanings.

Fodor suggests three main reasons to support the language of thought hypothesis. First, problems as to how human beings can learn to speak are more easily tackled if we postulate that the mind is essentially an internal representational system with linguistic structure. Second, the language of thought hypothesis brilliantly explains the productivity of human verbal behavior. In fact, from a finite set of words, human beings are in theory able to produce an infinite variety of propositions. This property of language requires that the representational content of propositions could be constructed from a finite set of atomic representations, which is exactly what the language of thought hypothesis states. Third, human thought is systematic, that is, certain thoughts are associated with certain other thoughts in such a way that any thinker who can understand one thought is also able to understand the others. For instance, every English speaker who can understand the meaning of "the cat is on the table and the mouse is under the chair" will also be able to understand the meaning of "the mouse is on the table and the cat is under the chair." This capacity of understanding seems to suggest an underlying implicit set of syntactical rules that govern the arrangement of the linguistic parts within the proposition, which is in accord to Fodor's hypothesis that at the root of human cognition there be a representational linguistic structure.

Fodor has subsequently tried to explain how the basic atomic representation can have meaning and content, as the computational theory of mind does not provide a

semantic account of how cognition works. Initially, Fodor endorsed a theory called *inferential role semantics*, according to which the content of a representation is in part determined by the inferential relations that it bears to other representations. For instance, “bachelor” derives its meaning from the inferential relation that it bears to “unmarried adult male” and vice versa, because both the concepts “bachelor” and “unmarried adult male” have the same internal semantic structure. However, Fodor has then argued against this view, because it would lead to semantic holism, a philosophical position which holds that any concept is potentially connected to any other concept. Fodor claims that semantic holism is incompatible with idea that concepts can be shared among people. In fact, since everybody practically holds some eccentric beliefs about everything, then if semantic holism were true, nobody could share any concept with anybody else.

Fodor therefore rejects all the theories that identify meanings and concepts in terms of both their inferential relations and internal structure and currently defends a view that he calls *informational atomism*. According to informational atomism, concepts are unstructured atoms whose contents and meanings are determined by certain informational connections with environmental phenomena. Thus, to possess the concept of “cat” is not to have certain beliefs about cats but rather to have an internal symbolic representation which is in the appropriate mind-world connection with the property of *being a cat*. The origin of these atomic concepts is, according to Fodor, innate (they are in fact supported by neural mechanisms), because in order to construct a mental representation of something one must already be equipped with a set of primitive concepts. However, although concepts are not *learned* on the basis of experience, they are nonetheless *triggered* by it.

According to Fodor’s view, acquiring a concept is not a matter of acquiring some beliefs but rather the other way round: beliefs are constructs out of concepts. Fodor therefore suggests that cognition has a compositional nature: conceptual basic elements form more complex representations, which in turn are embedded in propositional attitudes (such as beliefs, desires, hopes, etc.) having different mental shapes compared with their constituents. An astonishing example of how different elements can create a new meaningful shape is provided by the art of sixteenth-century Italian painter Giuseppe Arcimboldo, who was renowned for creating imaginative portraits entirely made of objects, such as fruits, flowers, fishes, etc. Figure 7.1 shows one of his most famous paintings, *Flora*, a portrait of a lady exclusively composed of flowers.

Fodor’s view of cognition as a computational process leads him to claim that the mind itself is largely composed of different modules, each of which can perform a different task. The idea that the mind has a modular architecture has been widely influential in cognitive science and has gained success in other areas of neuroscience as well, such as neuropsychology. According to Fodor, modules are characterized by specific features; in particular, they are *informationally encapsulated*, *domain specific*, *mandatory*, and *fast*. They are informationally encapsulated because they do not have access to all the information that the system elaborates but

Fig. 7.1 Giuseppe Arcimboldo (1527–1593), *Flora*, oil on panel (circa 1591), private collection, Paris, France (Image cropped by authors)



only to the inputs which are relevant for their task. They are domain specific because their activity is bound to certain types of representation (i.e., visual, auditory, sensory, etc.). Finally, they are mandatory and fast because they are automatically and quickly triggered by the stimuli to which they apply.

Cognitive neuropsychology has adopted among its assumptions a more refined concept of the modularity of mind, according to which modules are *isolable functional subsystems*, which can operate independently from other subsystems, even though not as efficiently as they could operate when the others are sustaining their activity. Neuropsychological syndromes can therefore be interpreted as the result of damage or disruption to one or more of these functional subsystems. Contrary to Fodor's description, this view considers modules as neither rigid nor computationally autonomous but incorporated in neuronal circuits, patterns, or configurations, which are able to maintain a degree of plasticity, so that under particular circumstances the brain is able to reorganize itself by replacing the function of the damaged module or subsystem with the compensation of the others.

A debated issue in neuroscience is whether the modular structure can be a global feature of the mind or just a feature of the peripheral representational systems. It can be argued that a modular organization seems to be very useful at the low level

of perceptions and sensations, where the perceptual scene can be parceled in separate and easily identifiable features, but not at the higher level of abstract thought, where information needs to be elaborated recursively by widespread neuronal networks able to form a global workspace, which in turn can sustain in parallel the integration of different cognitive functions (such as memory, language, reasoning, planning, decision-making, self-reflection, etc.). On the other hand, it has been suggested that even some higher-level functions can be fractionized in domain-specific modules. For instance, British psychologist Alan Baddeley (2007) holds that working memory be composed of four elements: a central executive and three temporary storage systems (a visuospatial sketchpad, a phonological buffer, and an “episodic buffer”).

Given the rigid function and computational autonomy of modules, Fodor claims that modularity exclusively characterizes the peripheral or sensory/input systems. By contrast, human higher-order thought (which is not strictly based on computational processes, as it is shown by intuitive and analogical reasoning) would prove that the central systems must be unencapsulated in order to borrow and use information coming from different perceptual and cognitive domains. This could imply that Fodor’s mental representational theory provides a convincing account of how the fringes of the mind work, but leaves its inner core as well as consciousness somewhat unexplained. In Fodor’s own words, “as things now stand, [...] consciousness looks to be the ultimate mystery about the mind” (2000, p. 99).

Essential Bibliography

- Fodor J (1983) *The modularity of mind*. MIT Press, Cambridge, MA.
In this book Fodor puts forward the hypothesis that the mind is largely composed of different modules, each of which is specialized in elaborating a certain kind of information in a certain way. This idea has gained popularity among the scientific community and, in its more refined versions, has been widely influential in several areas of neuroscience.
- Fodor J (1994) *The elm and the expert: mentalese and its semantics*. MIT Press, Cambridge, MA.
This book, which collects Fodor’s Jean Nicod Lectures, lively discusses both the idea that mental phenomena are computational representations and the language of thought (or mentalese) hypothesis. Fodor pays special attention to semantic issues, especially with regard to the relationship between mind and meaning.
- Fodor J (2008) *LOT 2: The language of thought revisited*. Oxford University Press, Oxford.
Fodor further develops his hypothesis about the language of thought as well as the idea that the nature of mind possesses a widespread compositional architecture. He also argues in favor of an atomistic theory of thoughts, which is – according to Fodor – the most useful model to account for the great variety of human behavior.

The (Almost) Truth of Physicalism

I believe that there is reason to think that intentional/cognitive properties are functionalizable. However, I am with those who believe that phenomenal properties of consciousness are not functional properties.
(Physicalism or Something Near Enough)

Jaegwon Kim is a well-known Korean American philosopher born on 12 September 1934 in Daegu (now in South Korea). His main philosophical interests are in the fields of philosophy of mind, metaphysics, and epistemology. Kim attended 2 years of college in Seoul, South Korea, as a French literature major, before entering Dartmouth College in 1955. At Dartmouth, he changed to a combined major in French, mathematics, and philosophy. Afterwards, he moved to Princeton University, from which he gained his PhD in philosophy. Since 1987, he has been the William Herbert Perry Faunce Professor of Philosophy at Brown University. In 1988–1989, he was president of the American Philosophical Association, Central Division. He also taught at Swarthmore College, Cornell University, University of Notre Dame, Johns Hopkins University, and the University of Michigan, Ann Arbor.

Kim's philosophical style is characterized by exemplar clarity and painstaking attention to analytical argumentation. His main concern is to unravel the knot of the mind-body problem within the framework of physicalism, the doctrine claiming that in nature only physical entities are causally relevant. In this context, the question whether the mind can bring about changes in the physical world appears to be crucial. As we have seen in Chap. 6, this problem had already been raised in 1643 by Princess Elizabeth of Bohemia in a letter to René Descartes, the father of modern philosophy. Specifically, Elizabeth formulated her question in the following way: "How the mind of a human being can determine the bodily spirits [i.e., the fluids in the nerves, muscles, etc.] in producing voluntary actions, being only a thinking substance. For it appears that all determination of movement is produced by the pushing of the thing being moved, by the manner in which it is pushed by that which moves it, or else by the qualification and figure of the surface of the latter. Contact is required for the first two conditions, and extension for the third. [But] you entirely exclude the latter from the notion you have of the body, and the former seems incompatible with an immaterial things" (Elizabeth to Descartes, May 1643; quotation taken from Daniel Garber, *Descartes Embodied*, Cambridge University Press, 2001, p. 172).

Elizabeth pointed out one of the main problems with Descartes' philosophy of mind, that is, how was it possible for an immaterial and unextended substance (mind) to have contact with a material and extended substance (body). This issue is known in modern philosophy of mind as the problem of *mental causation*, and in Kim's contemporary terms, the problem presents as follows: How can the mind have causal powers in a world that is fundamentally physical? Or, in other words, can the psychological features of mental states have causal influence on physical objects and events? Moreover, the problem of mental causation is intricately linked to another one: How can there be room for consciousness in a physical world, that is, a world made of material elements, which behave in accordance with physical laws?

On the one hand, the world we live in seems to consist of physical or material entities only; on the other hand, there seem to be good reasons for thinking that our behavior depends on cognitive processes that involve causal relations with elements of the outside environment. The reality of mental causation seems to be at the root of the possibility of psychology as a science capable to propose law-based explanation of human actions. And if a phenomenon is to play an explanatory role, its presence or absence must make a *causal* difference. Thus, if we want to maintain some sort of interaction between mental and physical entities, we should be inclined to concede that mind itself has a physical nature. This was the solution suggested to Descartes by Princess Elizabeth of Bohemia, who wrote in the same letter quoted above: "And I admit that it would be easier for me to concede matter and extension to the mind than it would be for me to concede the capacity to move a body and be moved by one to an immaterial thing". This is also Kim's strategy, which argues that the only way to make sense of mental causation is to *physicalize* the mind, which in turn leads to the idea that mental phenomena can to some extent be physically accountable.

In order to account for the relationship between mind and matter, Kim endorses the concept of *mind-body supervenience*, which he considers as a shared minimum commitment for all physicalist positions. According to the mind-body supervenience thesis, our mental life is totally dependent on bodily processes. Mind-body supervenience can be explained in terms of generalization from the subvenient (the neural substrate) to the supervenient (the mental state). Thus, whenever an individual is in the mental state M, then the individual must be in the physical state P, and every individual who has P has also M. In other words, physically indiscernible systems cannot differ in respect of their mental properties. Kim claims that this general view of mind-body supervenience lies at the core of any form of reductive physicalism. It is also a commitment shared by both functionalism and emergentism. However, as this concept seems to suggest a double ontology – by dividing natural phenomena into the subvenient physical states on the one hand, and the supervenient mental properties on the other – its endorsement within a monistic view of nature risks to cause more confusion than benefit.

Kim points out that the physicalist stance on the world commits us to accepting two principles: (1) the causal closure of the physical domain and (2) the causal exclusion. The former claims that "If a physical event has a cause at t, then it has a

physical cause at t .” The latter claims that “If an event e has a sufficient cause c at t , no event at t distinct from c can be a cause of e .” Following the exclusion principle, the sufficient cause c of e at time t may be either physical or mental. However, this instance is ruled out if the principle of exclusion is linked to the principle of the physical causal closure and the event e is identified with an event whose nature is purely physical. Consequently, the two principles joined together hold that only physical events can cause other physical events. Conversely, if we accept that event e is simultaneously caused by both physical and mental events, we would have to face the serious problem of *causal overdetermination*, given that the same event e would be causally overdetermined.

Therefore, Kim claims that if we want to save mental causation, we have to be prepared to take reductionism seriously. Otherwise both the physical causal closure and the exclusion principle compel us to consider mental phenomena as merely epiphenomenal, that is, superfluous by-products of brain activity, just like the shadows which objects cast on the ground are not material parts of those objects. This view was already supported by Thomas Henry Huxley (1884) who compared consciousness to the steam whistle of a locomotive. Huxley was perplexed as to the exact function of consciousness, to the point that he argued consciousness played no role in behavioral mechanisms: he argued that consciousness can neither cause nor modify animal behavior just like the steam whistle of a locomotive cannot influence the work of the locomotive’s engine.

Such epiphenomenalist viewpoint implies that a rigorous examination of human actions should deny the reality of consciousness, and thereby of all mental states correlated with this phenomenon. In fact, since epiphenomenalism upholds that the conscious mind is not part of the physical causal world, the entire removal of mental events from nature would not determine any change in the world as we know it, as long as all the physical properties of natural events are preserved. This implies that, given the physical causal closure of the universe, conscious mental events cannot interact with the physical reality in any way. But if we think of ourselves as consciously acting agents, the epiphenomenalist claim sounds deeply counterintuitive.

If we want to avoid epiphenomenalism and save mental causation, we have to be committed to what Kim calls a *conditional thesis*: “Mental phenomena must be physically reducible, if we want them to have a causal influence on the physical world.” However, Kim points out that in accepting this thesis, we are not compelled to think that the mind as a whole must be either all reducible or all irreducible. In fact, it could be that part of the mental domain is reducible, whereas the rest is not. Kim’s suggestion is that the intentional and cognitive aspects of mind are physically reducible, while the conscious phenomenal aspects are not. In other words, the psychological states that play a functional role in producing behavior – such as belief, intention, and desire – are susceptible to be functionally reducible, that is to say, they can be defined in terms of their causal roles. In fact, behavior would not have any meaning in the absence of these psychological drives. We are simply compelled to consider behavior as the product of cognition. According to Kim, this constitutes



Fig. 8.1 Peter Paul Rubens (1577–1640), *The Rainbow Landscape* (circa 1636), oil on oak panel, The Wallace Collection, London, England (Image cropped by authors)

a strong reason to interpret the intentional and cognitive mental states as describable in terms of their roles in behavior causation.

Contrary to intentional and cognitive mental states, the subjective and phenomenal features of the mind, often referred to as *qualia* (see Chap. 1), cannot be defined in terms of their functional roles. In fact, it seems that we do not feel the same commitment in interpreting the behavior as the product of qualitative conscious experiences. Kim is led to this point because he accepts the conclusions that can be drawn from the thought experiment of the inverted spectrum. It is possible to imagine people who behave like us despite their inverted spectra: what we perceive as blue is perceived by them as red and so on for all the other wavelengths – in Fig. 8.1, for instance, the colors of the rainbow should be inverted for these people. According to this argument, from the fact that we can imagine people different from us because of their inverted spectra but with the same behavior as us, we can infer that the qualitative aspects of mind do not have any role in behavior causation. Thus, the metaphysical hypothesis of qualia inversion is, according to Kim, the indication that qualia are not physically or functionally reducible. This is why he claims that physicalism is “almost” or partially true.

However, it can be argued that the thought experiment of the inverted spectrum is not logically consistent. At what point exactly should the inversion occur? It cannot occur at the level of light wavelength reflected by objects; otherwise, we would all have color qualia inversion, which means that the inversion would have no effect whatsoever. Similarly, it cannot occur at the level of the retina, where there

are three types of color receptors (cones): even if we changed the type of cone that is supposed to respond to blue into the type of cone that is supposed to respond to red, the pattern of their joint activation would remain the same. Intuitively it is to no avail to replace the coalitions of neurons that interpret stimulus R (red wavelength) coming from the eyes as “red” with the coalitions of neurons that interpret stimulus B (blue wavelength) coming from the eyes as “blue” at this level. In fact, it is not important what kind of coalition will do the job. What it is important, instead, is that at least two coalitions of neurons will activate in the brain in order for us to experience R as red and B as blue, respectively. To argue that a coalition could see R as blue and another one could see B as red presupposes that there can be extra elements that have to be added to the neuronal processes and change from one coalition to the other, so as to make the qualia inversion possible. However, the existence of these extra ingredients is rejected by the monistic ontology of physicalism. Thus, it seems that the inverted-spectrum argument might be fatally vicious, because it claims to prove its implicit presupposition, which is that the same neuronal process can realize either a red *quale* of a blue *quale*.

Essential Bibliography

- Kim J (1993) *Supervenience and mind*. Cambridge University Press, Cambridge. The book presents a detailed analysis of the concept of supervenience and the problem of mental causation in a clear and engaging style. The principles of causal closure of the physical world and the principle of causal exclusion are analyzed in the light of the concept of supervenience of mental states over physical (brain) states.
- Kim J (1998) *Mind in a physical world*. MIT Press, Cambridge, MA. In this book, the concept of supervenience is further discussed and considered as pivotal in the relationship between the mental and physical realms. The author critically appraises other philosophical positions on the mind-body problem, including Davidson’s anomalous monism (see Chap. 4).
- Kim J (2005) *Physicalism or something near enough*. Princeton University Press, Princeton. The book examines the themes put forward by Kim’s preceding works and reaches the conclusion that among all the theories, physicalism is the one which is nearest to the truth, even though consciousness is not susceptible to be thoroughly understood in physical terms.

The Inner Sense Theory of Consciousness

...the mind has no special properties that are not exhausted by its representational properties, along with or in combination with the functional organization of its components.
(Consciousness and Experience)

William G. Lycan (born on 26 September 1945) is an American philosopher who has made relevant contributions to a range of philosophical fields, including philosophy of mind, philosophy of language, epistemology, and metaphysics. Lycan earned his MA in 1967 and PhD in 1970, both from the University of Chicago. He taught at the Ohio State University until 1982, when he joined the University of North Carolina at Chapel Hill (UNC), where he is the current William Rand Kenan, Jr. Distinguished Professor. Lycan has also a co-appointment at the University of Connecticut.

Lycan explicitly offers a line of attack or programmatic assault to the problem of consciousness from a mere physicalist point of view. He in fact defends an entirely physicalist or materialist approach to the problem of consciousness, which according to him is not to be thought of as a single problem but rather as an intricate cluster of different questions, encompassing both empirical and theoretical ones. Lycan suggests that each question can become more tractable only once it has been carefully distinguished from the others. His strategy is then *to divide and conquer* (from the Latin motto “divide et impera”), namely, to separate different aspects of the conscious mind – which can be individually tackled by either empirical or theoretical tools. The final solution to the puzzle of consciousness can be reached only after putting the individual answers together. In fact, according to Lycan, consciousness is not only the problem of intentionality (i.e., mental aboutness; see Chap. 3) or the problem of qualia (i.e., the qualitative feature of experience; see Chap. 1) or the problem of what it is like to have a subjective viewpoint on the world (see Chap. 11), but all these problems altogether, as well as others.

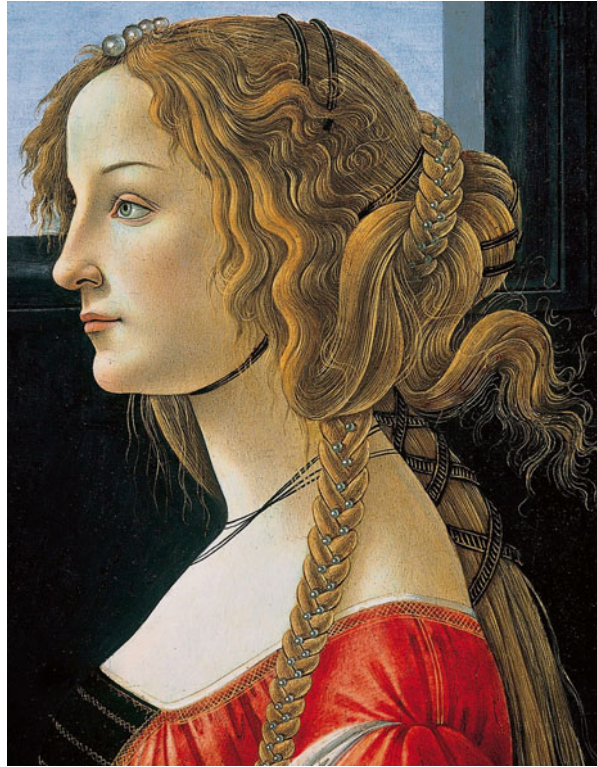
As a starting point, Lycan suggests that, before attempting any solution to the problem of consciousness as such, we must have a clear sense of what is meant by the terms *conscious* and *consciousness*. He therefore pinpoints the following eight uses of these words, each of which in turn identifies a different type of consciousness:

1. *Organism consciousness*. A conscious being is opposed to a nonconscious being if and only if the conscious being can think, sense, and feel.
2. *Control consciousness*. A conscious being is opposed to a nonconscious being if and only if the conscious being is awake, endowed with mental states and in control of its own actions in a way which is consistent with its mental states. This kind of consciousness is similar to Rosenthal's concept of *creature consciousness* (see Chap. 14) and to Ned Block's concept of *access consciousness* (see Chap. 26).
3. *Consciousness of*. A conscious being is aware of this or that. A conscious state is directed to something, which can be an external object, an abstract entity, a feel, etc.
4. *State/event consciousness*. A conscious state or event is opposed to a nonconscious state or event if and only if the subject is aware of being in that state or knows that that event is occurring within him or her.
5. *Reportability*. In a sense, one is conscious only of the things he or she can verbally report on.
6. *Introspective consciousness*. There is introspective consciousness when the individual focuses the attention on the internal character of his or her experience. This type of consciousness is a special case of state/event consciousness.
7. *Subjective consciousness*. This is the type of consciousness that can be described in the first-person point of view and directly refers to what it is like to be in a certain mental state.
8. *Self-consciousness*. This type of consciousness refers to having a sense of oneself as an individual who is distinct both from other individuals and the environment.

Lycan analytically examines each type of consciousness and concludes that the first two types (*organism consciousness* and *control consciousness*) are strictly associated with the mind-body problem as it is classically conceived (i.e., the problem as to how the mind can exert influence and control over the body). In turn, the third and the fourth types of consciousness (*consciousness of* and *state/event consciousness*) appear to be special cases of the problem of intentionality (i.e., the problem as to how mental states can be *about* or *directed toward* something). The fifth type (*reportability*) does not constitute a real philosophical problem, as it can be completely explained in empirical terms according to our linguistic faculties. The sixth type can also be considered as an empirical issue, along with the issue concerning the nature of different modes of perception and attention mechanisms. The seventh type, on the contrary, is considered by Lycan to be the real problem of consciousness within the philosophical framework of physicalism or materialism. The eighth type is not directly addressed by Lycan, but we could consider it, following his line of reasoning, as a special case of type three (*consciousness of*), in which the self is the object toward mental states are directed.

According to Lycan, subjective consciousness is in turn formed by different conceptual aspects, just as a rope is composed of different threads or a braid of hair,

Fig. 9.1 Sandro Botticelli (1445–1510), *Portrait of a Young Woman* (circa 1476–1480), tempera on wood panel, Gemäldegalerie, Berlin, Germany (Image cropped by authors)



as it is beautifully shown by Italian Renaissance artist Botticelli's *Portrait of a Young Woman* (Fig. 9.1). Subjective consciousness can therefore be further divided in different subproblems, which Lycan identifies in the following ones: the subject/object distinction (i.e., being in a particular mental state versus observing someone else's brain when he or she is in that mental state); the individual's immediate or privileged access to his or her mental states; the temporal anomalies between observationally detectable events and the individual's awareness of these events; the problem as to how and why consciousness has evolved; the epistemological problem of the inverted spectrum (i.e., the possibility that two identical persons can inversely perceive colors – what is red for one is blue for the other and so on; see Chaps. 9 and 13 for a discussion); and the possibility of philosophical zombies (i.e., creatures identical to us in every physical detail but with no conscious experience; see Chap. 1 for a discussion).

Lycan considers the issues regarding the subject/object distinction and the subject's privileged access to his or her mind as false theoretical problems, since they are simple facts which derive from the way things naturally are. In turn, the issue about the temporal anomalies between the actual events and the subject's awareness of

these events is an empirical question, which therefore is to be subsumed under the more general label of the *binding problem* (i.e., the problem as to how different brain patterns which codify for different aspects of a scene can merge together in order to produce a consistent and unitary picture of the world; see Chap. 17). Similarly, the issue regarding the role played by consciousness in the evolutionary process is to be thought of as an empirical question. On the other hand, the other issues are merely philosophical but, according to Lycan, they do not constitute the main conceptual difficulties for our understanding of conscious phenomena.

Lycan claims that all major conceptual problems concern the phenomenal nature of conscious experience, which is intrinsically characterized as being internal, immediately accessible from the first-person point of view, ineffable, perspectival, inaccessible to scientific descriptions expressed in third-person terms, and thereby, in a sense, inexplicable by science. Thus, according to Lycan, *qualia* are distinctive, introspectible, and monadic features of conscious experience, such as the particular shade of green of a patch of lawn that one sees, or the specific pitch of sound that one hears. This characterization, however, raises a number of conceptual difficulties, including the impossibility of a thorough account of mental states in the scientific objective way – which, in turn, leads to the so-called problem of the *explanatory gap* (i.e., the impression that the correlation or correspondence between certain neuronal patterns with certain conscious experiences – and not with others – requires a further explanation).

Despite recognizing the existence of *qualia* and thereby the intrinsic subjective nature of conscious experience, Lycan argues that the conscious mind has no special features that cannot be explained by virtue of its representational properties and of the functional organization of its components. He thus suggests a representational approach in order to understand the qualitative nature of subjective or phenomenal consciousness. In this view, *qualia* are thought to be intentional and representational contents of sensory states. The difference between the characters of the phenomenal experience (whether it is visual, auditory, etc.) depends on the causal role played by the representations within the functional organization of the mind. But according to Lycan, this is not enough to realize the phenomenal awareness of perceptions, as it is possible to have sensory experiences unaccompanied by consciousness, such as unfelt pains or unnoticed visual sensations (the neuropsychological condition of blindsight is a paradigmatic example of this occurrence).

Lycan therefore proposes a further functional processing that adds awareness to experience, so that experience itself can become the object or the content of a higher-order mental representation. As a result, consciousness appears to be a sort of *inner sense*, which is able to subjectively perceive sensory experiences. This inner sense theory applies to every mental state (regardless of whether it has a qualitative nature or not) experienced by the subject. This theory might be considered as being equivalent to Rosenthal's higher-order thought theory of consciousness (see Chap. 14); there is however a crucial difference between the two approaches. Lycan's inner sense processing is a kind of self-monitoring, a sort of proprioception based on an internal attention mechanism, which is more *perception-like* rather than *thought-like*. Thus, Lycan calls *higher-order perception* (HOP) the

representation whose content is another representation. Furthermore, he argues that, compared to higher-order thoughts (HOTs), HOPs have the advantage of being able to allow nonconceptual representations and, thereby, a noncognitive account of conscious phenomena. Higher-order perceptions result in unique and distinctive modes of presentation, thus emphasizing the first-person perspective of phenomenal consciousness.

Lycan's inner sense theory of consciousness seems to effectively address some of the criticisms that can be raised against Rosenthal's view (see Chap. 14 for a discussion). In fact, consciousness does not appear to be simply a cognitive process, but rather an inner sensing device which does not entirely rely on the possession of certain concepts. Furthermore, contrary to HOTs, HOPs are not susceptible to be regarded as epiphenomena, since perceptions are by definition causal processes, whose influence on behavior is based on the degree of the stimulation. However, similarly to Rosenthal's model, Lycan's theory might not be easily translated into neurobiological terms. In fact, it is unclear what neuronal mechanisms are supposed to be involved in the generation of higher-order perceptions. As we shall see in the section dedicated to the neuroscientific theories of consciousness, neuroscience seems to suggest that conscious experience may rely upon widespread associative processes throughout the brain rather than on a further processing level in which information is represented in a perception-like fashion.

Essential Bibliography

- Lycan W (1987) *Consciousness*. MIT Press, Cambridge, MA.
This book engagingly discusses different philosophical accounts of the nature and origin of consciousness and attempts to coherently accommodate what Lycan regards as the most valuable aspects of each view. Lycan defends an original position that he calls "homuncular functionalism," according to which our brains are functionally organized information-processing systems.
- Lycan W (1996) *Consciousness and experience*. MIT Press, Cambridge, MA.
In this book, Lycan further develops his functionalist approach to consciousness with a painstaking analysis of the philosophical arguments against a purely physicalist view of the mind. He also argues that consciousness is a cluster of intricate problems, which can only be tackled successfully if they are considered as separate issues.

The Mysterious Flame of Consciousness

*It is almost as if we have been designed to be struck
by a problem that we are constitutionally unable to solve:
the very self-consciousness that makes us aware
of the problem is (part of) what prevents us from solving it,
because of the concepts that are generated by such self-consciousness.*
(Consciousness and its Objects)

Colin McGinn (born on 10 March 1950) is a British philosopher whose work has mainly focused on metaphysics and philosophy of mind. McGinn gained a BA (Hons) and an MA in Psychology, in 1971 and 1972, respectively, both from the University of Manchester. In 1974, he received a BPhil from the University of Oxford. McGinn taught at University College London and from 1985 to 1990 held the position of Wilde Reader in Mental Philosophy at the University of Oxford. He was visiting professor at several universities, including the University of California, Los Angeles, Rutgers University, City University of New York, and Princeton University. In 1990, he became full professor in the Philosophy Department at Rutgers University, thus working alongside American philosopher and cognitive scientist Jerry Fodor (see Chap. 7). In 2006, he joined the University of Miami, where until 2013 he held the position of Professor of Philosophy and Cooper Fellow.

McGinn is well known in the community of philosophers of mind for claiming that the nature of consciousness is far beyond our understanding. According to him, neither the direct or conceptual examination of consciousness (through introspection and the painstaking philosophical analysis of concepts) nor the scientific study of the brain (through experiments and neuroimaging investigations) can ever explain the mechanisms or identify the real causes that bring about our conscious experiences. American philosopher Owen Flanagan has dubbed this theoretical position “new mysterianism.” In contrast to the old mysterianism, which could be associated with classic dualism (see Chap. 6), new mysterianism does not claim that the conscious mind is something mysterious because it has incomprehensible supernatural aspects, but more simply because it eludes our epistemological skills. Thus, the first doctrine is based on ontological grounds (consciousness is supernatural), while the second one is based on epistemological grounds (consciousness cannot be

understood with the help of our philosophical and scientific methodologies). In light of that, the so-called hard problem of consciousness (see Chap. 1) will remain totally inaccessible to human reason.

According to McGinn, we are in an epistemological stalemate: we know that there is a genuine mind-body problem, but at the same time, we do not possess the conceptual and empirical tools to solve it. What is more, there is the impression that our knowledge of the mind-body problem outruns our abilities to formulate it correctly. It is as if we can intuit more about this problem than we can explicitly say. McGinn comes to this conclusion by relying on the distinction between knowledge by acquaintance and knowledge by description, a distinction that was formulated by the great British philosopher Bertrand Russell. To have knowledge by acquaintance means that we can have access to something without any intermediary process of inference. This kind of knowledge is therefore direct and immediate. Conversely, to have knowledge by description means that we can know something because of a line of reasoning which can be turned into words. This kind of knowledge is therefore indirect and mediate, specifically conveyed by linguistic terms. The two processes of knowledge are independent of each other, because it is possible to be directly aware of something without being able to describe it, just like it is possible to describe something without having any direct personal experience of it (think about an art critic who describes the painting technique of an artist, who has instead a direct acquaintance with his or her style that the critic lacks).

McGinn claims that we are able to know consciousness by acquaintance only. On the one hand, we are directly aware of our own consciousness, but, on the other, we cannot say what it really is. We cannot provide a propositional explanation of consciousness, although we are constantly acquainted with conscious phenomena. In other words, despite having immediate access to what it is like, for instance, to see the blue sky or to drink red wine, we cannot thoroughly account for what constitutes these experiences. This is why we often resort to various metaphors to speak about consciousness, such as a mirror, a stream, a spotlight, a theater, etc. However, all these metaphors do not grasp, according to McGinn, the real essence of conscious experience. McGinn therefore concludes that the propositional knowledge given in physical terms is inevitably defective when we try to understand consciousness.

It is worth noting how Australian philosopher Frank Jackson comes to a similar conclusion, albeit through a route which seems to be opposite to McGinn's. Jackson is famous within the field of philosophy of mind for having proposed a thought experiment aiming to prove that reductive physicalism is an inadequate theory to account for our conscious experience. He suggests to imagine a neuroscientist, called Mary, who has spent all her life in a black and white room, where she has conducted research on the brain mechanisms of color vision. Mary has a deep physical knowledge of what colors are; in fact, she is able to give a thorough propositional account of how the vision of color depends on the interaction within the human eye between light receptors and electromagnetic radiations of certain

wavelengths. However, because of her reclusion in the black and white room, Mary has never had any conscious experience of what it feels like to see colors.

This scenario seems to be quite the opposite to the one illustrated by McGinn. Hypothetically, Mary has an ideal knowledge by description of conscious visual phenomena (because of her life spent studying the neural mechanisms of color vision), but she lacks any direct knowledge by acquaintance about what it is like to see all the spectrum of colors (because she has always lived in a black and white room). In his intriguing thought experiment (*Gedankenexperiment* in philosophical jargon), Jackson asks his readers to imagine what would happen if Mary could finally go out of her black and white room and actually see the world. At this point, two alternatives might occur: In the first one, looking at the sky, Mary would comment “I already knew that blue would look like that!”; in the second one, she would be surprised and report her newly acquired knowledge “So this is how blue looks like!”. Intuitively, everyone tends to think that the second alternative is much more likely than the first one, rather, that the first alternative is utterly implausible, if not absurd. In fact, no one can ever be able to feel something by listening to or by reading descriptions, no matter how precisely a sensation can be described. Jackson therefore draws the conclusion that the reductive version of physicalism cannot ever provide a thorough or complete explanation of consciousness.

Jackson’s thought experiment about Mary the neuroscientist has been extensively debated. Here it suffices to say that it might not provide the ultimate proof that reductive physicalism is false, but rather a strong argument for the distinction between events which are experienced in the first-person perspective and events which are experienced in the third-person perspective. In fact, one of the interpretations of Jackson’s thought experiment is that what Mary actually learns, by seeing for the first time colors, is not an ontological truth (about what there really is out there) but an epistemological one (about what we know about what there is out there). As a result, physicalism might be true from the ontological point of view (consciousness is a brain process), but might also enshroud a dual epistemological nature (we can know what it is like to be conscious only in the first-person perspective), and thereby be unable to provide a reductive account of consciousness.

McGinn seems to offer a further argument for this case, provided that Russell’s distinction between knowledge by acquaintance and knowledge by description is to be accepted (within the philosophical literature there are diverging opinions about this). Since consciousness appears to be acquaintance based, we can only have an *implicit* knowledge of it, whereas any form of *explicit* account of this phenomenon is beyond our imagination. In light of that, every theoretical attempt to conceptualize consciousness is destined to miss the mark. According to McGinn, even the concept of intentionality is neither necessary nor sufficient for explaining consciousness (see Chap. 3 for a different position about intentionality). And the same can be said of neuroscience. McGinn claims that any complete neuroscientific theory of conscious experience should be able to demonstrate that conscious states are to be reduced to brain processes, as it historically happened in the case of temperature (which was found to be identical to the mean kinetic energy of molecules) or water

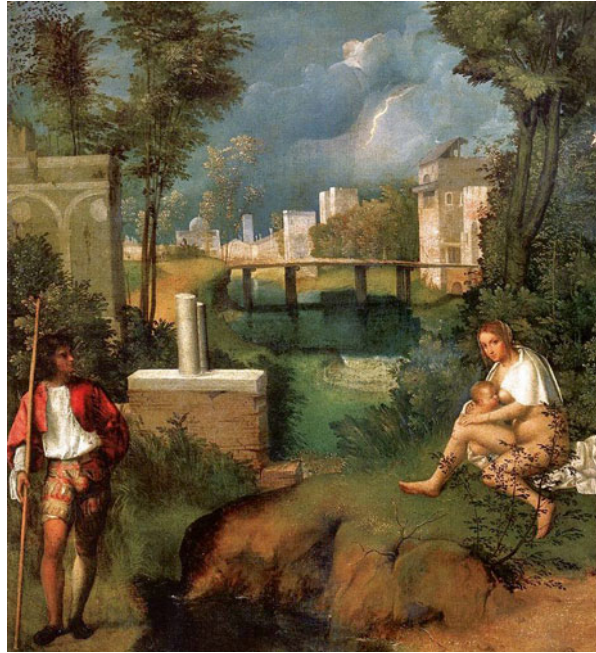
(which was found to be identical to H₂O). A simple statement of *correlation* between brain processes and conscious states would not be sufficient, because it would not provide a genuine reduction of mind to matter.

In spite of his belief that consciousness is a natural phenomenon (and therefore amenable to scientific exploration), McGinn questions that such scientific theory will ever be found. According to him, there is no known property of the brain which can simply be considered as identical to consciousness. What is more, mental phenomena seem to have a different conceptual status compared to the entities studied by the physical science. As an example, McGinn examines the concept of space, which seems to be fundamental in our knowledge of the physical world but, at the same time, seems to be of less (if any) use in understanding mental phenomena. McGinn endorses Descartes' intuition that the conscious mind is unextended in space (see Chap. 6). Similarly to Descartes, McGinn argues that spatial concepts are inadequate to describe what consciousness is. Contrary to neuronal structures, conscious experiences cannot be described as having an exact location, size, shape, parts, etc. When we locate mental events within the brain, this simply is, according to McGinn, "a sort of courtesy location." He therefore suggests that in order to reconcile our intuition of the nonspatiality of consciousness with the scientific investigation of nature, we would need a completely new concept of space, deriving from a full *perspective shift* rather than just a *paradigm shift*. However, this is exactly what human beings are not able to do, unless they evolved to become different cognitive beings altogether.

At first blush, McGinn's philosophical arguments seem appealing, but it could be argued that they lead to a number of questionable consequences. In fact, if it is true that consciousness is an acquaintance-based phenomenon, how can we be sure that we have the same acquaintance of the same phenomenon? Why does McGinn claim that conscious experiences are nonetheless natural phenomena instead of assuming, for instance, that they are supernatural events? Why should his acquaintance of consciousness be more deeply grounded than the one proposed by Descartes (see Chap. 6), who thought that consciousness was an immaterial entity? Why does McGinn trust Descartes on the nonspatiality of consciousness but not on the immaterial nature of consciousness? In the light of McGinn's view, consciousness seems to fall prey to any sort of different interpretations. The plight of consciousness would be similar to that of *The Tempest*, a famous painting by Italian Renaissance artist Giorgione (Fig. 10.1), whose real meaning remains thus far mysterious.

As we shall see in the following chapters dedicated to the neuroscientific theories of consciousness, the neuroscientific investigation of consciousness has already yielded a few preliminary results, which are promising and raise intriguing suggestions. For instance, there is common consensus among neuroscientists that consciousness is a phenomenon occurring in time and in a serial fashion, and most neuroscientist would agree that it depends on the functional connections between specific brain structures. These characteristics are also strictly linked to the spatial organization of the brain. It would appear that there is no possibility to understand the former without the latter. The scientific endeavor has just begun, and, undoubtedly, the future is ripe with surprising discoveries.

Fig. 10.1 Giorgio Barbarelli da Castelfranco, known as Giorgione (circa 1477/8–1510), *The Tempest* (circa 1508), oil on canvas, Gallerie dell'Accademia, Venice, Italy (Image cropped by authors)



Essential Bibliography

- McGinn C (1991) *The problems of consciousness*. Blackwell, Oxford.
In this book, McGinn introduces the reader to the idea that consciousness, despite being a natural phenomenon, is far beyond the reach of human intellectual faculties. Human minds are simply not equipped to understand consciousness, neither through the empirical instruments of scientific investigation nor through introspection.
- McGinn C (1999) *The mysterious flame: conscious minds in a material world*. Basic Books, New York.
This book further develops the thesis that human mind cannot unravel the mystery of consciousness. This provocative position is discussed in detail and raises intriguing questions about the nature of mind and brain as well as the meaning of dreams and self-reflection.
- McGinn C (2004) *Consciousness and its objects*. Oxford University Press, Oxford.
This is a collection of some of the most significant essays written by McGinn on the mind-body problem and the nature of consciousness. The distinction between knowledge by acquaintance and knowledge by description is introduced and applied to conscious phenomena. McGinn also proposes that the intrinsic nature of consciousness might be nonspatial.

What Is It Like to Be a Conscious Mind?

*An organism has conscious mental states
if and only if there is something
that it is like to be that organism –
something it is like for the organism.
(What Is It Like to Be a Bat?)*

Thomas Nagel is an American philosopher currently teaching Philosophy and Law at New York University. In 1974, he published an article – titled *What Is It Like to Be a Bat?* – that had enormous resonance in the community of philosophers of mind. The expression used by Nagel in his article has become a mesmerizing phrase in the common usage among philosophers to indicate the subjective feature of any phenomenal experience, that is, the what-is-it-like-to-be-a-something. This issue would have been called by another philosopher, David Chalmers, the *hard problem* of consciousness studies (see Chap. 1).

The tenet of Nagel’s thought is that the subjective point of view is *intrinsic* in what it means to be a conscious mind and therefore irreducible to an objective explanation of human nature given in mathematical and physical terms. In fact, being a conscious mind is to have a subjective perspective on the world just like as having a phenomenal subjective point of view is tantamount to be in a conscious state. In a sense, the relationship between consciousness and subjectivity is so tight that we cannot have one without the other. But, importantly, a subjective point of view should not be confused with the sense of self. Indeed, a sense of self implies a cognitive elaboration of feelings, while to have a subjective perspective means just to have a privileged, exclusive, and inviolable access to the world.

The intrinsic privacy and exclusiveness of one’s conscious point of view eliminates the possibility of applying a conceptual reduction in order to account for consciousness in an objective way. In fact, in his article Nagel clearly claims that the subjective character of experience cannot ever be captured by any reductive analysis of the mind, since its absence is logically compatible with all the physicalist accounts of mental phenomena. We find again the argument of conceivability that we have seen when we dealt with Chalmers’ property dualism. Given that it is conceivable that robots or automata behave like humans without being conscious, then it is supposed that there is no logical necessity to ascribe consciousness to mere physical or functional processes. In other words, physicalism does not exhaust the analysis of conscious experience.



Fig. 11.1 Henry Singleton (1766–1839), *Ariel on a Bat's Back* (1819), oil on canvas, Tate Britain, London, England (Image cropped by authors)

Still, according to Nagel, there is another more compelling reason for being suspicious of the explanatory power of physicalism. In fact, if physical accounts of mental reality were true, then the very phenomenological character of mental reality should be described in physical terms. At a closer look, however, we discover that such an achievement is impossible. The reason is that every objective physical theory of mind would neglect the subjective conscious experience which is essentially connected with a particular point of view. In order to better explain this issue, Nagel put forward an argument that has become one of the most famous in the philosophy of mind. The argument is based on the assumption that bats have experience. Thus, there must be something that it is like to be a bat. Bats largely perceive the world by echolocation, that is, by detecting the reflections from objects of their high-frequency shrieks. The brain of a bat is able to make discriminations of distance, shape, size, and motion of things by elaborating the information coming from echoes. It seems, therefore, that the bat's perception of the external environment is radically different from ours.

Despite this difference, however, Nagel asks whether we can understand the inner life of a bat in some way. Although echolocation is by no means similar to any form of human perception, we can perhaps figure out what it is like to be a bat by the power of imagination (Fig. 11.1). With our mind's eye we could try to realize what it feels hunting insects at dusk and hanging upside down from a cave ceiling. Still, Nagel claims that imagination cannot help us understand the bat's

point of view. In fact, as hard as we can try, we will never be able to be aware of the inner life of a bat, since imagining this sort of experience tells us only what it would be like for *us* to behave as a bat behaves. The conclusion is that the bat's subjective point of view – as well as the subjective point of view of any other creature – will forever remain completely elusive. In other words, there are facts in the world (i.e., subjective phenomena experienced by creatures different from us) which are not susceptible to being described by human language. Thus, we should accept the existence of such facts, even though we are not able to verbalize or comprehend them.

This line of reasoning, however, could lead to counterintuitive consequences. In fact, if the only point of view that we can grasp is ours, then we are destined to fall into a sort of radical subjectivism, as any attempt to understand the point of view of a different organism would be in vain. What is more, even the points of views of all the other similar organisms which belong to the same species would be out of reach. Every subjective standpoint would be a token, a unique and unrepeatable event in time and space. The step to a solipsism of sensations, therefore, might be short. Nevertheless, Nagel is able to recognize and avoid the problem. His solution is to favor a sort of *type subjectivism* or *objective phenomenology*. In other words, according to Nagel, it is possible for an organism to comprehend what the quality of another organism's experience is, provided that they are sufficiently similar and can thereby adopt the other's point of view. Therefore, at least within the same species, there is a sense in which phenomenological features of an experience are objective, and it is possible to understand the subjective point of view of others. In addition, there is nothing to rule out the possibility that organisms which belong to different species can somehow understand each other, as long as there is not an unbridgeable phylogenetic gap between them. For instance, every person can comprehend what it is like to have a visual perception of a rainbow, while a dog or a Martian scientist cannot, even though they would be able to understand the human concepts of colors and rainbows and their relationship in the language of physics.

This ingenious argumentation notwithstanding, theoretical problems remain. A questionable point is what Nagel intends when he speaks about "sufficient similarity" between organisms. At which point exactly an organism becomes dissimilar to another, and in what terms this fact affects the capacity of adopting the other's perspective? Another problem is that the subjective character of an experience should not be conflated with the acceptance of a different point of view. In theory, someone can perfectly adopt and comprehend the subjective standpoint of another person or sentient creature without being able to simultaneously feel the subjective quality that that person or creature feels in having his/her/its subjective point of view. In light of the foregoing considerations, it is arguable how an objective phenomenology might ever be possible.

On the basis of the arguments presented so far, one can expect that Nagel claims that physicalism has failed. But this is not the case. Although it remains a mystery why there must be something that it is like to be certain physical processes – as long as these physical processes are to be thought identical with certain mental processes – Nagel suggests that, after all, a kind of nonreductive physicalism might

be true. In *The Psychophysical Nexus*, his second important essay dedicated to the philosophy of mind, Nagel attempts to grapple with this conundrum.

In this essay, Nagel tries to propose a sort of noncontingent psychophysical identity theory. Although the subjective character of conscious experience will never be accounted for by a physical and/or functional description of the world, Nagel considers the relationship between mental and physical phenomena as necessary. This position is in deep contrast to Kripke's analysis of mental properties (*Naming and Necessity*, Harvard University Press, 1980), which claims the contingency of this correlation. As it is possible to conceive a world in which mental states exist without brains, and vice versa, Kripke argues that the connections between the two must be contingent. Nagel's reply is that the contingency is only conceptual and a priori, so that we cannot exclude that there might be the case of an a posteriori necessary relation (i.e., based on empirical facts) between mental and physical processes. It is interesting to note that this kind of response rules out the possibility of philosophical zombies, that is, sentient creatures physically identical to human beings but with no conscious experience. In fact, if conscious mental states must necessarily be physical phenomena, then it is not possible that a creature with the latter does not also have the former. Moreover, Nagel's scenario has another noteworthy result: if robots will ever be able to acquire consciousness, they will acquire it in a different type from ours.

Most of the essay is thus dedicated to understanding how the necessary relationship between mental and physical events might hold, so that the physiology of the brain, the phenomenology of sensations, and the manifest behavior can be knotted in a single nexus. In order to achieve this result, Nagel proposes to accept three truths: (1) a conceptual and contingent truth that every mental state plays a specific functional role in relation to behavior, (2) a conceptual and necessary truth that every conscious mental state has its particular phenomenological features, and (3) a nonconceptual and necessary truth that every conscious mental state has its particular physiological properties. Nagel claims, therefore, that there is a single event to which we can refer in two ways, one via the mental description which conveys the subjective character of conscious experience and another via the physical description which conveys the relevant patterns of brain physiology.

The inescapable implication of this theoretical position is that neither the mental nor the physical concepts can ever afford a thorough understanding of conscious phenomena. Both of these conceptual views are limited. In reaction to this limitedness, Nagel's last book on the mind-body problem *Mind and Cosmos* is a further attempt to broaden the theoretical outlook on consciousness. The thesis of the book is that the basis for psychophysical identity is something more basic than what can be grasped by purely mental or physical accounts. Therefore, the dual aspect of reality would derive from the fact that this primordial component appears to be both objectively physical from outside and subjectively mental from inside. The basic constituent of the universe, in other words, should have properties that explain both its physical and mental character. Moreover, Nagel suggests that this essential element could pervade the whole cosmos. Thus, the explanation of consciousness would extend beyond the biological framework so as to involve the entire natural order.

In a sense, Nagel proposes a fairly radical revision of evolutionary theory. In fact, he argues that, by virtue of the protomental element of reality, the likelihood of the emergence of consciousness must have been latent in nature. The process of evolution, therefore, would not be driven solely by blind physical causes but also by teleological propensities. In a nutshell, Nagel suggests that we could live in a world with a predisposition to the development of consciousness.

This idea is fascinating but, at the same time, highly problematic. Do we really need a mysterious concept, such as a universal protopsychic constituent, in order to explain another allegedly mysterious phenomenon like consciousness? It can be argued that the mystery remains, even though it is brought to another level. What is more, to reintroduce teleology in physics implies being in disagreement with its reformation inaugurated by modern scientists in the seventeenth century, which abandoned the Aristotelian tradition that accepted as true that things have a natural tendency toward ends or goals. This is not to say that teleology is an ill concept tout court but that it should find, if valid, a new mathematical formulation in order to make it coherent with the edifice of the modern science.

Essential Bibliography

- Nagel T (1979) What is it like to be a bat? *Philos Rev* LXXXIII:435–450; reprinted in: Nagel T (1979) *Mortal questions*, Cambridge University Press, Cambridge.
This paper presents an intriguing line of reasoning in order to sustain that the subjective character of any experience is intrinsic and nonreducible to objective physical terms. In virtue of its thought-provoking quality, it has become one of the most famous and discussed articles in the philosophy of mind.
- Nagel T (2000) The psychophysical nexus. In: Boghossian P, Peacocke C (eds) *New essays on the a priori*, Clarendon Press, Oxford; reprinted in Nagel T (2002) *Concealment and exposure and other essays*, Oxford University Press, Oxford.
This essay provides a painstaking inquiry into the possibility that the relationship between mental and physical aspects of the world can be necessary rather than contingent. Accordingly, the author argues for the challenging idea that mind and brain might be knotted in an inextricable nexus.
- Nagel T (2012) *Mind and cosmos*. Oxford University Press, Oxford.
This is the most recent essay published by Nagel on the mind-body problem and probably his most ambitious one. The author aims to give a global picture of both mental and physical realities by arguing that these two concepts describe phenomena which are the manifestations of a more elementary underlying component.

Consciousness As an Enactive Perceptual Device

*[Experience is] not caused by and realized in the brain,
although it depends causally on the brain.
Experience is realized in the active life of the skilful animal.
(Action in Perception)*

Alva Noë (born in 1964) is a philosopher whose activity has mainly developed within the fields of the philosophy of mind and perception. He earned his BPhil from the University of Oxford and his PhD from Harvard University. In 2003, he joined the University of California, Berkeley, as an Associate Professor and as a member of the Institute for Cognitive and Brain Sciences. From 2011 to 2012, he was Distinguished Professor of Philosophy at the City University of New York Graduate Center. He was also visiting scholar or fellow in several universities, including the Institut Jean Nicod in Paris and the Oxford Center for Cognitive Neuroscience. He is currently Professor of Philosophy at the University of California, Berkeley.

Part of Alva Noë's work has been devoted to the development of a model to explain the mechanisms of perceptual consciousness, that is, our state of awareness when we are engaged in the perception of objects. Perceptual consciousness is strictly associated with phenomenal consciousness, the qualitative awareness of things, although the two concepts are to be thought of as independent. This is exemplified by dreams (when there is phenomenal awareness but not perceptual consciousness) or by rare neuropsychological conditions, such as blindsight (when there can be a sort of perceptual consciousness that something has been seen without the accompanying phenomenal awareness of what precisely has been perceived).

According to Noë, perceptual consciousness is to be considered as having an intrinsic enactive nature. Consciousness and perception are in fact the results of the interaction between environmental stimuli and the sensory systems that perceive them. This model relies on a sensorimotor approach to consciousness and perception and has been gaining interest within the neuroscientific community (see Chap. 26), along with the idea that cognition appears to be largely determined by the body's structure and way of functioning (hence the label *embodied cognition*). Furthermore, the sensorimotor theory attempts to dispose of the concept of highly detailed mental representations of the external world (a model that depicts

the mind as the *mirror of nature*) in favor of gestalt images, which are afforded by the environment as soon as interaction occurs.

According to the sensorimotor theory, conscious perceptions are not the mere results of different firing patterns of the cerebral cortex, but of a mode of the whole body acting in the world. Accordingly, the way things look, sound, smell, or feel depends on the complex mode by which the body actively explores its environment. Perceptual contents are not simply produced by the brain, but rather enacted through the skilled sensorimotor activity of the body. Accordingly, Noë argues that conscious experience is not fully realized and caused by the brain, although it necessarily depends on its functional organization. Rather, experience emerges from the active life of a skillful organism.

With regard to vision, in order to explain how the brain, the body, and the environment can shape a perceptual content, Noë suggests the metaphor of a painter. Seeing is, according to him, very similar to painting. Just as painting is an ongoing process during which the painter's eyes alternate from the scene to the canvas and vice versa, so our eyes probe the world in cyclic and active explorations that deliver a new partial cognitive uptake of information each time. In other words, just as the production of a picture does not only result in the figure depicted on the canvas but rather in the dynamic engagement among the painter, the scene, and all the tools used for painting – as we can paradigmatically see in Vermeer's *The Art of Painting* (Fig. 12.1) – so our conscious visual contents are not represented within the brain, but instead interactively carved out of a cyclic world-engaging process.

Although this kind of interactive process is similar for every type of perception, the engagement of our sense organs differs according to sensorimotor dependencies. These dependencies or contingencies are relations between movements or changes and sensory stimulation. Although such relations are distinctive for each sensory modality, they all are the result of a loop or cycle which connects the perceived objects with fluctuating patterns of sensory stimulation. These sensory patterns can be caused by the perceiver's movements (as when the head and the eyes are moved to scan a visual scene) or by the movements of the object itself or finally by other elements of the environment (as changes in illumination or in light source). Noë argues that through the implicit knowledge of the sensorimotor dependencies, we are able to *enact* (i.e., actively realize) a perceptual experience.

In light of this enactive model of perception, both the content and the character of our perceptual experiences – regardless of whether they are visual, auditory, tactile, etc. – are explained by the particular sensorimotor dependencies that are involved in the loop between the world and the perceiver. Noë's enactive model of perception thus provides an alternative account of *qualia* (the phenomenal qualities of experience; see Chaps. 1, 9, 11, and 26): qualia are in fact no longer thought to be intrinsic and ineffable aspects of sensations but rather *distinctive sensorimotor profiles* which are engaged in the act of perception. In other words, it is not the presence of specific connatural features which gives to experience its characteristic phenomenal flavor, but the presence or absence of particular loops of sensorimotor actions linking the world to the perceiver's body and his or her sensory cognitive patterns.

Fig. 12.1 Johannes Vermeer (1632–1675), *The Art of Painting* (circa 1666), oil on canvas, Kunsthistorisches Museum, Vienna, Austria (Image cropped by authors)



Noë's enactive theory of perceptual consciousness has significant consequences with regard to some debated philosophical issues. For instance, the sensorimotor model ensures that it is impossible to conceive of philosophical zombies (see Chap. 1), because it implies that if two perceivers have exactly the same sensorimotor skills and discriminatory capacities, then they must also have the same perceptual experience. In fact, the content and character of any conscious perceptual experience could be recreated insofar as the same body of sensorimotor skills is rebuilt. What is more, the sensorimotor framework can reconcile the concept of a reality as mind independent with the concept of a world as perceived from the subjective viewpoint of an embodied agent. This in turn leads to the consequence that creatures with different sensorimotor skills are necessarily confined to different "perceptual worlds." In fact, creatures endowed with other "sensorimotor tunings" would not be able to directly experience the world in which the other ones live.

The enactive model of perception can also elegantly explain the possibility of neuronal rewiring, namely, the case in which a brain area, normally involved in a specific perceptual activity (for instance, hearing), can become, thanks to an early reconnection, part of a sensorimotor loop which is characteristic of another perceptual

activity (for instance, vision). As it has been shown by experiments on ferrets, if appropriately embedded in the visual dynamic cycle, part of the auditory cortex can take on visual functions.

The concept of sensorimotor dependencies can equally explain similar results obtained in experiments on human beings, in which the visual system is temporarily bypassed with technological devices able to substitute visual inputs with tactile or auditory stimulation. In the case of auditory-visual substitution system, the visual inputs coming to a head-mounted camera are translated into auditory stimuli. Different locations of objects are indicated by different pitches in sound, so that objects which are located in the high part of the visual field are codified with high-pitch sounds, whereas objects which are located in the low part of the visual field are codified with low-pitch sounds, and so forth. Experiments show that participants can distinguish different kinds of objects, such as plants, circles, statues, crosses, etc., because they are able to learn the auditory signature patterns (i.e., the sensorimotor dependencies) which are distinctive of objects.

Noë's enactive approach radically broadens our perspectives on consciousness. Rather than passively receiving the world as a bunch of data which are to be extrapolated and represented within the brain, the conscious mind actively probes the world, makes contact with it and gradually gains it through inquiry and exploration. The world is therefore reached rather than simply given to the mind. However, important this aspect might be for an accomplished science of consciousness, the sensorimotor dependencies could only account for the instance where a higher level of conscious functional organization is achieved rather than being the essential element of the contents of consciousness. In fact, none of the arguments put forward by the sensorimotor advocates seem to support the radical conclusion that phenomenal consciousness is not caused by or realized in the brain. On the contrary, neuroscience has been providing evidence suggesting that certain perceptual skills and experiences are brought about by the activity of specific brain circuitries. Thus, perceptual experience could be linked to specific forms of neuronal processing that appear to be largely insensitive to sensorimotor cycles. For instance, experiments show that the grip size is entirely determined by the true size of the target object, even though the conscious vision of the object is altered by illusion – as it happens in the “Titchener circles illusion,” in which subjects misjudge the relative size of a disk because it is surrounded by a ring of larger or smaller circles.

These experiments strongly suggest that the process underlying visual awareness may operate quite independently of those underlying the visual motor control of actions. In other words, perceptual consciousness might depend on specific brain processes that have little to do with fine-detailed sensorimotor mechanisms and much more to do with representations that assign inputs to categories, types, and locations of things, so as to use them for cognitive reasoning. In light of that, the sensorimotor loops would be the typical channels through which neuronal patterns are activated and mental representations recruited in order to construct a fully fledged conscious picture of the world.

Essential Bibliography

- Noë A (2004) *Action in Perception*. MIT Press, Cambridge, MA.
The book introduces the reader to the thought-provoking idea that perception is not something which passively happens to us, but something that we actively do. Noë proposes that perceptual consciousness is a skillful activity of the body as a whole, which deeply relies on the interaction among the body, the brain, and the environment.
- Noë A (2009) *Out of our heads: why you are not your brain, and other lessons from the biology of consciousness*. Hill and Wang, New York.
In this book, Noë explores the implications of his enactive approach to perception for the neuroscientific understanding of the conscious mind. He contends the common neuroscientific idea that consciousness is a functional activity confined within the brain and argues in favor of a broader perspective which considers consciousness as an active process encompassing both the body and its environment.
- Noë A (2012) *Varieties of Presence*. Harvard University Press, Cambridge, MA.
The book provides a painstaking philosophical argumentation about how the world is present to our minds. Noë intriguingly suggests that the world is gradually probed and gained rather than simply achieved. The world in itself has no meaning, apart from the degree of our engagement with it. The astonishing consequence of this approach is that education and skill acquisition can radically change the way we construct the world and transform our own consciousness.

The Rise and Fall of Functionalism

...whatever the program of the brain may be, it must be physically possible, though not necessarily feasible, to produce something with the same program but quite a different physical and chemical constitution.
(Philosophy and Our Mental Life)

Mental states cannot literally be “programs”, because physically possible systems may be in the same mental state while having unlike “programs.”
(Representation and Reality)

Hilary Putnam (born on 31 July 1926 in Chicago, Illinois, US) is one of the most important contemporary philosophers. He has been a central figure in analytic philosophy since the 1960s, as he provided key contributions to several fields of philosophy, ranging from philosophy of mind, language, and mathematics to ethics and metaphysics. After having studied at the University of Pennsylvania and Harvard University, he gained his PhD in 1951 from the University of California, Los Angeles, with Hans Reichenbach and Rudolf Carnap as his supervisors. In 1965, after a long residence at the Massachusetts Institute of Technology, he moved to Harvard where he taught Mathematical Logic. Currently, he is Cogan University Professor Emeritus at Harvard University.

Hilary Putnam is famous for his willingness to apply as much the same degree of scrutiny and criticism to his own theoretical ideas as to those of others. He has therefore endorsed different philosophical positions across time. With regard to philosophy of mind, he initially supported the view that mental processes are functionally accountable in computational terms. Putnam’s version of functionalism gained immediate success and still is one of the most popular outlooks on the nature of mind among philosophers and cognitive scientists.

Functionalism holds that mental processes (such as pain, beliefs, desires, etc.) can be defined by their causal relations to other mental processes, internal and external stimuli, and behavioral responses. Thus, a person’s desire of drinking a glass of water is not only related to his/her internal sense of thirst and a certain activation of neuronal networks but also to moving his/her hand in the direction of the glass, as well as to the beliefs that there is a glass of water on the table nearby and that the only way to quench the feeling of thirst is to drink water, and so forth. Mental states are therefore identified with their functional role within a *functional organization*, namely, the network formed by a set of interrelated functional states. In this view, the human mind is a specific kind of functional organization.

The key concept of functionalism is called by Putnam as *functional isomorphism*. According to this principle, two systems are functionally isomorphic if there is a correspondence between the states of one and the states of the other that preserves functional relations. In other words, if system 1 and system 2 are functionally isomorphic with each other and in system 1 state A is always followed by state B, then also in system 2, state A must always be followed by state B.

The growing consensus that has characterized functionalism is mainly due to the fact that this approach is strictly associated with cognitive science, a field of research that underwent a great expansion during the last decades and is still very active. Cognitive science developed as a reaction to behaviorism, a doctrine according to which the existence of mind is an unnecessary postulate, as human beings can be understood in terms of behavioral patterns only, that is, by using simple stimulus and response schemes. In contrast, cognitive science claims that mental states are computational processes of the brain, which elaborate inputs (stimuli) in order to produce outputs (responses).

Within this debate, Putnam significantly contributed to the idea of considering the mind as a computational device. He argued that the traditional mind-body problem is linguistic and logical in its nature and can be dissolved if we think of the brain in analogy with a computing machine. In fact, mentality could be compared with the internal logical operations of a computer. Just as we can describe computers in terms of both their structure (hardware) and their programs (software), so we can describe human beings in terms of both their physical structures and their mental processes. It seems therefore that there is a significant analogy between humans and computers. In fact, both can be accounted for in terms of physical states ruled by physical laws, as well as in terms of logical operations (computers) or mental processes (humans) ruled by laws of reasoning.

The idea of comparing brain activity to the functional organization of machines has a long history. The metaphor of hydraulic pipelines was used by the ancient Greeks to describe how the vital fluids could flow into the nerves. In the seventeenth century, British philosopher Thomas Hobbes thought of ideas and associations as being minute mechanical motions inside the head, similar to clockwork mechanisms, while in the nineteenth century, German scientist Hermann von Helmholtz conceived of the nervous system as a telegraph. Finally, before the advent of the computer, twentieth-century British neurologist Charles Sherrington poetically described the brain as an “enchanted loom” that constantly weaves our mental lives (Fig. 13.1).

Regardless of metaphors, all these images rely on the idea that it is possible to establish equivalence between mental activity and the functional organization of computing machines. This idea has led Putnam and many other thinkers to promote the thesis that mental states are *multiply realizable*. This thesis claims that every type of mental state can in principle be realized in different ways, in human brains, as well as in other animal brains or even in non-biological structures. In other words, the same type of mental event, say “pain,” can be instantiated by different kinds of physical processes. Pain and all the other mental events are therefore considered as functional states of a whole organism, so that to know that an organism is in pain



Fig. 13.1 Vincent van Gogh (1853–1890), *A Weaver's Cottage* (1884), oil on canvas mounted on panel, Museum Boijmans Van Beuningen, Rotterdam, Netherlands (Image cropped by authors)

involves knowing something about the causal roles played by that mental state in the functional organization of this organism. What is more, this functional organization can be expressed in logical and mathematical terms, which have no reference to the physical substrate of the organism. Putnam took this point to the extreme, by claiming that we could all be made of cheese and it would not matter, provided our functional organization is maintained.

Functionalism is a kind of identity theory of mind, as it identifies mental states with certain physical processes. In the case of human beings, for instance, mental states are supposed to be equivalent to certain brain processes, even though both states and processes can be described differently. Still, by virtue of the principle of multiple realization, functionalism is also consistent with a nonreductive approach to the mind-body problem. Since different occurrences of the same mental state can be functionally realized by different types of physical states, there is no strict correspondence between certain types of mental states and certain types of physical states. In fact, Putnam pointed out that whatever the program of the brain may be, it should be essentially possible to produce something with the same program but with an entirely different physical and chemical constitution.

In developing his philosophical reflection, Putnam subsequently changed his position regarding the computational nature of mental states to the extent that he now considers functionalism wrong in identifying mental states with computational processes. There are four main objections raised by Putnam against functionalism. The first objection is a version of the inverted-spectrum argument (see Chap. 8 for a discussion of this argument), according to which it is possible to imagine two persons having an identical functional organization but different color perceptions. In other words, person A would see red as red and blue as blue, while person B would see red as blue and blue as red, even though internally the functional role of their mental states has not changed.

The second objection to functionalism asks us to imagine the possibility that two individuals, say Peter and Mary, can share the same belief, such as “water is wet,” even though they possess different functional organizations. In other words, the same mental state (water is wet) can be part of a functional organization (Peter’s mind) in which has causal relations with certain mental states (water is made of H_2O) and, at the same time, be part of another functional organization (Mary’s mind) in which has causal relations with certain other mental states (water is made of xyz). According to functionalism, we should be forced to conclude that this is not possible, because if two individuals have different functional organizations, they cannot share the same belief or mental state, which is very counterintuitive.

The third objection to functionalism introduces a thought experiment that can be considered as the reverse image of the one introduced by the second objection. Whereas in the second objection, we were asked to imagine two individuals with different functional organizations but the same belief, now Putnam asks us to imagine two individuals with the same functional organization but different beliefs. This thought experiment, which provides the foundation to the third objection to functionalism, is one of the most discussed arguments in the philosophical literature and aims to show that meanings cannot be identified completely with mental processes occurring in people’s heads, because mental contents are at least in part externally determined. This position is called “semantic externalism” in philosophical jargon. In brief, the thought experiment runs as follows. Oscar, an individual on Earth, has a doppelganger (i.e., a molecule-by-molecule identical copy) named Toscar on Twin Earth, an exact replica of our planet except for the fact that water is made of xyz rather than H_2O . All the beliefs held by Oscar about water on Earth refer to a substance made of H_2O , while all the beliefs held by Toscar about water on Twin Earth refer to a substance made of xyz. Thus, despite their identical functional organization, Oscar’s and Toscar’s thoughts about water have different meanings and contents, which is in direct contrast to functionalism, according to which mental contents are entirely determined in internal functional terms.

The fourth objection is directed at dismantling the principle that is at the very core of computational functionalism: the thesis of multiple realization. Putnam points out that the concept of multiple realization is unclear and has little theoretical power, because it allows that almost everything with the proper functional

organization can have a mind. If this is the case, then the difference between system A and system B, for instance, a human being and an automaton, cannot be accounted for in functional terms, but only in terms of inputs and outputs, or stimuli and behavior. As a result, functionalism collapses on behaviorism, because possessing a certain functional organization is equivalent to having certain behavioral dispositions.

In addition to Putnam's objections, functionalism has been criticized with other arguments. For instance, the functionalists' perspective is thought to be in trouble when dealing with qualia or subjective mental phenomena. The philosophical zombie argument allows us to conceive that a zombie could have the same functional organization of one person but with no experience of qualitative properties (see Chap. 1 for a discussion of this thought experiment). Another strong argument against the idea that the thesis of the multiple realization of mental states implies their irreducibility has been put forward by philosopher of mind Jaegwon Kim (see Chap. 8). Kim points out that higher-order mental properties could be realized in different structures that nonetheless share lower-order and reductive-base properties. For example, being at a certain temperature can be realized in different bodies made of different materials, but all the bodies' molecules must share the lower-order property of having a certain mean kinetic energy.

Apart from sophisticated philosophical argumentations, empirical considerations are at present probably the most serious challenge to functionalism. As already said at the end of Chap. 5, neuroscientific research is gathering a number of data showing that the relationship between brain structure and brain function has to be taken into account in order to explain the nature of mind. According to this empirical approach, which is called *embodied cognition*, mind is deeply rooted in bodily structure and brain architecture, suggesting thereby that functional computation is not the whole story.

Essential Bibliography

- Putnam H (1975) Mind, language, and reality. Philosophical papers, vol 2. Cambridge University Press, Cambridge.
The book is a collection of Putnam's essays covering different topics in the fields of philosophy of mind, language, and metaphysics. It includes two seminal essays for the computational version of functionalism: *Philosophy and Our Mental Life* and *The Nature of Mental States*, in which Putnam argues that the mind is simply the functional organization of a computational system.
- Putnam H (1988) Representation and Reality. The MIT Press, Cambridge, MA.
In this book, Putnam argues that the analogy between a computational system such as a computer and the mind cannot account for the complex nature of mental states, as well as the faculty of reasoning and language. Putnam claims that these problems are unsolvable if tackled in functional terms only.

- Putnam H (2001) *The threefold cord: mind, body, and world*. Columbia University Press, New York.

With a lucid and attractive prose, Putnam deals with two largely debated philosophical topics: the relationship between perceptions and reality, on the one hand, and between mind and body, on the other. In grappling with these vexing problems of philosophy, he put forward insightful suggestions, by relying on the work of three great philosophers of the past: William James, John L. Austin, and Ludwig Wittgenstein.

The Higher-Order Theory of Consciousness

Explaining what it is in virtue of which conscious states differ from mental states that aren't conscious is the principal goal of a theory of consciousness. And it's fairly straightforward to get a start on that question. When a mental state is conscious, the individual that's in that state is conscious of it; when a mental state fails to be conscious, that individual is in no way whatever conscious of that state.
(Consciousness and Mind)

Philosopher David Rosenthal (born in 1939) has been particularly active in the field of philosophy of mind and consciousness. He is the current Professor of Philosophy and Coordinator of the Graduate Center's Interdisciplinary Concentration in Cognitive Science at the City University of New York (CUNY). He received his education from the University of Chicago and Princeton University.

Rosenthal has proposed to consider consciousness as related to two different concepts: the first one, which he calls *creature consciousness*, roughly corresponds to the neurobiological meaning of being awake or alert, whereas the second one, which he calls *state consciousness*, roughly corresponds to what we speak of when we refer to the contents of consciousness. The first concept introduces a distinction between various degrees of consciousness and between creatures that are conscious and creatures that are not. The second concept introduces a distinction between mental states which are conscious and others that are not or between contents of thoughts that can enter consciousness and others that cannot, as well as cases in which the same content can become conscious or remains unconscious.

According to Rosenthal, the first type of consciousness (creature consciousness) can be accounted for in neurobiological terms, on the basis of the degree of brain activation. On the contrary, the second type of consciousness (state consciousness) constitutes the real problem for both philosophy and neuroscience. In fact, we will never be able to explain consciousness if we cannot clarify what distinguishes a conscious mental state from an unconscious mental state.

The solution put forward by Rosenthal is that a mental state is conscious if it is accompanied by a specific type of thought. Rosenthal observes that if someone is not at all aware of a mental state, this state is not a conscious one. In other words, given that for being conscious of something we must be in a conscious state that has for content the object which we are aware of, then being aware of a mental state is a necessary condition for that mental state to be a conscious one. Thus, in order

for a mental state to be conscious, it has to be the content of another mental state. In fact, there is no reason to suppose, Rosenthal points out, that we should be conscious of our mental states in a way which is different to the way we are conscious of everything else.

At first sight, this theoretical construction would seem destined to a vicious regress, because we could suppose that we should be conscious not only of one conscious state but also of that further second mental state that has for its content the mental state which we are aware of. In turn, we should be conscious of that further second mental state by virtue of another third mental state that has for its content the second mental state, and so forth. However, Rosenthal explains that in this case there is no circularity, because it is not required that a mental state must itself be conscious in order to make another mental state conscious. In other words, one thing is the individual's being conscious *of* something, and another thing is a *mental state's being conscious*. Rosenthal calls *transitive consciousness* the fact that an individual is conscious of something and, accordingly, *transitivity principle* the assertion that mental states are conscious only if an individual is somehow conscious of them. Our feeling of being conscious of something in a way that seems to us immediate is also explained by the fact that we do not need to be conscious of a second-order mental state in order to be conscious of something.

According to Rosenthal, the transitivity principle is the only criterion that helps distinguish between conscious and nonconscious mental states. And there is just one consistent way of being transitively conscious of things: we are conscious of something when we have an appropriate thought about it. Since thoughts about mental states are on a higher order with respect to their contents, Rosenthal calls these thoughts *higher-order thoughts* (HOTs), and his model about the conscious mind the *higher-order theory of consciousness*. Similarly to the wanderer depicted in a famous painting by Caspar David Friedrich, who stands above a sea of fog and sees some peaks getting out of the mist, HOTs look down to the unconscious sea of the mind and intentionally focus on mental states, thus making them conscious (Fig. 14.1).

Rosenthal claims that the HOT hypothesis is able to explain what is distinctive with regard to the phenomenon of introspection. In fact, since a mental state is conscious only if it is accompanied by a related HOT (which rarely is a conscious state itself), in the case of introspection, the HOT is in its turn conscious in virtue of another HOT of a superior level. Moreover, the HOT theory is able to account for the deep sense of unity of any conscious experience, which seems to always belong to the same self. This is so because a HOT does not only refer to a mental state but also to that state as belonging to somebody, that is, a HOT about something makes the individual conscious of it only when the HOT represents the mental state as being present to the individual's self. From the theoretical point of view, the referencing self could be a different mental construction for each HOT, but since there is no reason to subjectively feel different selves, we have the impression that the self does not change through time. It is as if there is the default assumption that every HOT always refers to the same individual. Thus, according to Rosenthal the HOT



Fig. 14.1 Caspar David Friedrich (1774–1840), *The Wanderer Above the Sea of Fog*, oil on canvas (1818), Kunsthalle Hamburg, Germany (Image cropped by authors)

model provides an explanation as to why we seem deeply involved in consciously feeling a unitary sense of self.

According to Rosenthal, the higher-order model of consciousness is supported by the reportability of mental states. In fact, a verbal expression that an individual is in a specific mental state presupposes a corresponding thought that the individual is in that state. In other words, the verbal report of a conscious mental state is just the expression of a HOT about that state. Reportability is therefore a reliable indicator that the reported state is conscious. However, this does not imply that creatures unable to report their mental states are not conscious. In fact, they can be conscious at

least to a certain degree, as there is vast empirical evidence that human infants and many other animals show the ability to think. Accordingly, these creatures should have HOTS about their mental states, even though these HOTS are not directly linked to language.

Rosenthal argues that the HOT hypothesis is also able to offer a plausible account of phenomenal consciousness. According to this theory, the particular way we are conscious of a qualitative experience depends on how the accompanying HOT conceptualizes the properties of that experience. In fact, HOTS make fine-grained discriminations between different phenomenal qualities possible. For instance, a gourmet can discriminate between different tastes much better than the average person, because he or she is able to express his or her gustatory sensations with a greater variety of concepts. This seems also true for any other field of human activity. No one can be fully conscious of seeing what a professional ice skater does if he or she does not hold the technical concepts of axel, lutz, salchow, etc. Thus, the fact that we need to learn new concepts for better discerning qualitative mental states is evidence, according to Rosenthal, that we are conscious of those phenomenal qualities in virtue of HOTS that embody such concepts. This suggests that being conscious of certain qualitative features is tantamount to having a specific HOTS for those experiences.

The HOT model seems to elegantly solve the hard problem of consciousness, that is, the *what it is like* for somebody to have a certain experience (see Chaps. 1 and 11). This presupposes, however, that all the phenomenal properties can be wholly accounted for in cognitive terms. But it may be argued that this is not the case. For instance, we could conceive of an amnesiac who has lost all the concepts linked to the color red. Even though the amnesiac is no longer able to express the color red in words (or simply think about it) if he or she is placed in front of a red wall, it seems difficult to sustain that the amnesiac does not *consciously perceive* the color red, provided that he or she does not have any visual organic impairment. In fact, it is reasonable to suppose that if the amnesiac is placed in front of a yellow wall, he or she will be able to say whether or not there is something different with respect to the red wall, even though he or she cannot tell what the difference consists of. Thus, *being able to recognize* that an object is red seems to be different from *being able to have a mere conscious sensation* of a red object. And that should hold true for any other qualitative feature.

In addition to theoretical difficulties, the HOT hypothesis may lead to consider consciousness as an epiphenomenon, that is, a secondary property of certain mental states which results in a by-product of the HOTS that are directed to those mental states. In effect, Rosenthal has claimed that consciousness has no significant function in human behavior. According to him, this does not imply that consciousness does not have causal impact at all on psychological processes, but only that its impact is too small and, thereby, insignificant to the benefit of the organism. Further work is needed to clarify how it is possible to sustain, from a physicalist point of view, a clear-cut division between “significant” and “insignificant” causes. Moreover, there are neuropsychological data showing that people who suffer from different types of impairment in conscious brain processes (for instance, involving

representations of space, body, movement, etc.) tend to no longer engage in behaviors related to those representations. Thus, it seems reasonable to hypothesize that if certain behaviors occur in association with specific conscious states, then those specific conscious states must play a *significant* causal role in the brain processes that result in specific behaviors.

Finally, from a neurobiological point of view, it is not clear what the neural correlates of HOTs are exactly supposed to be. Recent neuroscientific findings seem to suggest that consciousness is a widely distributed property throughout the cerebral cortex and that what distinguishes a mental state that is conscious from a mental state that is not conscious might be a variation in the intensity of activity of reentrant neuronal patterns, rather than a further separate level in the elaboration of information.

Essential Bibliography

- Rosenthal D (2005) *Consciousness and mind*. Clarendon Press, Oxford.
This book collects some of Rosenthal's most influential essays about the mind-body problem and the nature of consciousness. The higher-order thought model of consciousness is explained in detail, so that readers can understand how this theory could account for introspection and the experience of phenomenal qualities.
- Rosenthal D (2008) Consciousness and its function. *Neuropsychologia* 46:829–840.
In this article, Rosenthal reviews the findings of important neuroscientific studies in the light of the concept of consciousness provided by the higher-order thought model and claims that the conscious mind is not causally relevant, even though it could have a sort of minimal causal effect.

Biological Naturalism

Consciousness is an ordinary feature of certain biological systems, in the same way that photosynthesis, digestion, and lactation are ordinary features of biological systems.
(Why I Am Not a Property Dualist)

American philosopher John Searle (born on 31 July 1932 in Denver, Colorado) has made outstanding contributions to different fields of philosophy, including philosophy of language, social theory, and philosophy of mind. He studied as an undergraduate at the University of Wisconsin-Madison and became a Rhodes Scholar at Oxford University, from which he obtained a doctorate in philosophy. His first major contribution was in the field of philosophy of language, where he developed into a comprehensive theory the insightful idea of Oxford philosopher John Austin that linguistic utterances can be seen as *speech acts*, that is, things that we do with words. Subsequently, his interest for the semantic analysis of language led him to the study of the concept of “intentionality” (the property of mind to be about or directed to things and states of affairs in the world in order to represent them) and this, in turn, to the nature of mind and consciousness (see Chap. 3 for a thorough discussion of intentionality).

With regard to consciousness, Searle’s starting point is to critique both the dualist and materialist perspectives. According to Searle, both types of approach are based on true premises but lead to false conclusions. Dualists correctly emphasize the irreducibility of consciousness to physical terms; however, they incorrectly claim that conscious states are not ordinary parts of the physical world. On the other hand, materialists correctly insist that the universe is made of material things only, such as physical particles and fields of force; however, they incorrectly argue that irreducible states of consciousness do not exist. These two radically different claims can be simply accounted for by the fact that dualists and materialists have conflict opinions about our ordinary language, whose vocabulary appears to maintain a clear-cut conceptual dichotomy between mind and matter. In fact, dualists consider the ontological categories derived from our traditional way of speaking as real, while materialists think of them as deeply misleading. For the dualist, it is evident that the mind must be kept separate from the body – don’t we say that “we *have* brains” rather than “we *are* brains”? For the materialist, on the contrary, it is mistaken to consider mind and brain as separate entities, and who is inclined to maintain this distinction is misled by our use of language.

Searle agrees with the materialist that the partition between the mental and the physical is a confusing and obsolete way of thinking, but instead of embracing the materialist conclusion that mind is reducible to matter, he agrees with the dualist that consciousness has a dimension which cannot be reduced ontologically to physical elements. This dimension is made by the qualitative and subjective aspects of consciousness (we have already encountered this line of reasoning; see Chap. 11). The qualitative aspect of conscious states is what makes them unique, private, and different. The experience of tasting wine is different from listening to Beethoven's Moonlight sonata, and both are different from seeing a rainbow or smelling tobacco. Moreover, all these experiences are private and unique for every individual. There is, in fact, a peculiar feeling attached to every conscious experience, something that *it is like* to have that experience, which is only accessible to the individual who is having that experience.

Thus, the qualitative aspect of consciousness leads necessarily to its subjectivity. In fact, conscious states exist only if they are experienced by a subject, so that if there is no subjectivity, there is also no experience. To describe this essential feature of consciousness, Searle adopts the expression *first-person ontology* (here the term "ontology" is used to indicate a mode of existence), as opposed to the *third-person ontology* of atoms, molecules, and inanimate objects, which can exist independently of conscious creatures. This is the reason why Searle claims that consciousness cannot be explained by a physical account. In fact, since consciousness has a first-person ontology, it cannot be reduced to something that has a third-person ontology, something that can exist independently of experiences. However, Searle argues that consciousness remains a biological phenomenon caused by biological processes and capable therefore to interact with other biological processes. All mental phenomena, he says, find their cause in the functioning of the brain, in virtue of lower-level neuronal processes, which realize the mental states as higher-level system features.

Thus, from the biological point of view, consciousness is not dissimilar from other biological functions, like digestion, photosynthesis, and the secretion of bile. Just as the liver can produce bile, so the brain can originate consciousness. However, Searle claims that the fact that the mind can be accounted for in biological terms does not imply that consciousness is nothing but neuronal processes. To better explain this point, he introduces a subtle distinction between *causal reducibility* and *ontological reducibility*. The former entails a relationship of dependence and can be applied to consciousness, because consciousness can be causally accounted for by functional processes going on in the brain, whereas the latter entails a relationship of identity and cannot be applied to consciousness, because consciousness has a first-person mode of existence, which is different from a third-person ontology. But even if its ontology is different, consciousness is not a distinct and separate phenomenon, something over and above its neuronal underpinnings, but it is, rather, a functional state the brain can be in, just in the same way as liquidity and solidity are states that water is able to assume.

However, it could be argued that this sort of distinction between causal reducibility and ontological reducibility cannot be sustained. A materialist, who does not

endorse an eliminativist approach toward the mental linguistic category, could reply that the subjective flavor of conscious states is *a way of knowledge rather than a way of existence*. We have acquaintance with conscious phenomenal states in both ways: through neuronal firings patterns in the brain and through our personal experience. Ontologically speaking, however, the descriptions we deal with refer to one entity only, that is, a certain set of physical processes in the brain, so that the relationship between neuronal firing patterns and conscious experiences appears to be of perfect identity: the former are the latter and the latter are the former. Thus, the materialist reply to Searle is that we constantly deal with a double epistemology (here the term “epistemology” is used to indicate a mode of knowing) rather than a double ontology. Following this view, consciousness could be accounted for by a *first-person epistemology* (qualitative and subjective) on the one hand and a *third-person epistemology* (quantitative and objective) on the other. Within this framework, the analogy with water’s liquidity or solidity would fall short of genuinely expressing a valuable ontological difference. In fact, from the ontological point of view that physics gives us, there is no liquidity or solidity, but only atoms and chemical bonds among them, which are interpreted by our brains as sensations of liquidity and solidity. Thus, the metaphor would be tautological: to say that consciousness is like liquidity would be exactly the same as saying that consciousness is like the conscious experience of liquidity. The materialist could therefore strongly argue in favor of the idea that if we have a causal reduction, then we must also necessarily have an ontological reduction.

Searle calls his theory *biological naturalism*, according to which consciousness is thought to be nothing more than a biological process, even though a process of a very special kind. In addition to the qualitative and subjective features of consciousness, Searle highlights a third aspect which characterizes every conscious experience: conscious states are always realized as a unity. Our sensations and feelings come as ingredients of a unified conscious field. There are no separate sensory perceptions, but rather a consistent mixture of sensations (coming from vision, smell, taste, hearing, and touch) and proprioceptive impressions.

Searle distinguishes two possible approaches that neuroscience can have in order to deal with the unitary aspect of consciousness: the building block theory (which thus far has been the standard way of studying consciousness) and the unified field theory (which he is inclined to favor; see Chap. 25 for a similar proposal). According to the building block theory, the conscious field is a combination of small elements (each of which has a specific content) that coherently form the field. Within this framework, to understand how the brain can engender even one of these components is to explain the problem of consciousness. Scientific research adopting this approach therefore aims to find the neural correlates of consciousness (NCC), which are causally necessary and sufficient for the existence of a conscious experience (with regard to the quest for the NCC, see Chap. 17). According the other approach, the unified field theory, the field of consciousness must be activated in order to have even one single content of conscious experience, which therefore cannot be explained in isolation. In fact, every conscious experience is a particular modification of the field. On the grounds of this second approach, the main issue about

consciousness that needs solution is not what the NCCs are, but rather how a brain sensory system (say, vision) can introduce visual experiences into an already activated unified conscious field and, above all, how the brain is capable to create this unified conscious field in the first place.

John Searle is also a strenuous opponent of the idea that machines could replicate human consciousness. To demonstrate that this is in principle not possible, at least by the methods used by current programs of artificial intelligence, he put forward a thought experiment called the “Chinese room.” The argumentation aims at showing that the Turing test is wrong. This test was devised by Alan Turing, a brilliant mathematician and computer scientist, to demonstrate that a machine is as intelligent as a human being if we are not able to distinguish the former from the latter on the basis of the machine’s behavior. The test is conceived as follows. Suppose you talk online with someone by texting messages. You know that you can be talking either with a human being or with a machine. If you are not able to find out from the messages that you receive whether you are talking with a human being or a machine, then you have to admit that machines can replicate the human mind with all its features, including consciousness.

Searle’s reply is that machines apparently seem to understand language, but actually they do not. What machines can do, instead, is to combine different symbols following a predetermined set of rules. To explain this point, Searle invites us to imagine a room in which someone who knows English language only sits alone following a manual of instructions for manipulating strings of Chinese characters in reply to cards written in Chinese and delivered from outside (Fig. 15.1). This way, the people outside the room will be under the erroneous belief that the person inside the room really understands Chinese. Computers are in the same situations as the human being in the Chinese room: just as someone can manipulate Chinese characters without any understanding of their meaning, so computers can consistently combine linguistic symbols by merely using syntactic rules without understanding semantics and meanings.

The Chinese room argument has the merit to raise the important point of semantics: if a computer is supposed to replicate the human mind, then it must also understand the significance of what it does rather than meaninglessly mimicking human behavior. It is arguable, however, whether the argument could prove that in principle computers will never be able to have an intelligent conscious mind. The solution of this issue remains on the side of the empirical research and only future studies will shed some light on the fascinating issue of artificial consciousness.

Essential Bibliography

- Searle JR (1992) *The rediscovery of the mind*. MIT Press, Cambridge, MA.
In this pioneering book, Searle puts forward his main ideas on how to tackle the mind-body problem and the question of the origin and nature of consciousness. Adopting a crystal clear and analytical style, Searle argues for the distinction between first-person and third-person ontology and claims the centrality of consciousness for the study of the mind.

Fig. 15.1 The opening verses of *The Tao Te Ching*



- Searle JR (2002) *Consciousness and language*. Cambridge University Press, Cambridge.
This book hosts an interesting collection of essays regarding themes in the fields of philosophy of mind, psychology, and philosophy of language. Searle is able to weave different conceptual threads in order to give a coherent picture of human beings as conscious, free, and rational agents.
- Searle JR (2004) *Mind: a brief introduction*. Oxford University Press, Oxford.
The book is a clear introduction to the main debates and theories in the philosophy of mind. Searle lucidly examines the strengths and weaknesses of the most important theoretical approaches aiming at solving the mind-body problem and concludes that consciousness has to be considered as a biological phenomenon susceptible of being accounted for by science.

Part II

Scientific Theories of Consciousness

The Global Workspace Theory

A theatre combines very limited events taking place on stage with a vast audience, just as consciousness involves limited information that creates access to a vast number of unconscious sources of knowledge. Consciousness seems to be the publicity organ of the brain. It is a facility for accessing, disseminating and exchanging information, and for exercising global coordination and control.
(In the Theater of Consciousness)

Bernard J. Baars (born on 1946 in Amsterdam, Netherlands) is an American neuroscientist whose work has fundamentally contributed to make consciousness a respectable subject of scientific inquiry. Baars received a BSc in Psychology and a PhD in Cognitive Psychology from the University of California, Los Angeles (UCLA). In 1977, he was appointed as Professor of Psychology at the State University of New York (SUNY) and from 1986 to 2000 was Institute Faculty Professor of The Wright Institute in Berkeley, California. In 2000, he became Senior Fellow in Theoretical Neurobiology at The Neurosciences Institute, San Diego, California, where he is currently Affiliate Research Fellow. Baars is founding coeditor of *Consciousness and Cognition: An International Journal* and founding editor of *Science and Consciousness Review*. He is also founding president of the *Association for Scientific Study of Consciousness* (ASSC).

Baars has developed one of the most successful theories in consciousness studies. According to his model, consciousness is like an information gateway to the brain, because it allows a widespread structure of neuronal networks to operate in order to integrate, provide access, and coordinate the processing of many specialized brain sites, which would otherwise operate autonomously. This widespread architecture of neuronal networks has been described by Baars using the metaphor of the *global workspace theater*, a sort of cognitive stage in which mental functioning occurs, at both conscious and unconscious level. Within this picture, consciousness would be like a bright spot on the stage of working memory (i.e., the part of short-term memory which is concerned with immediate perceptual and linguistic processing), guided by selective attention.

In Baars' view, consciousness is intimately correlated with the global workspace, though not identical to it. There are in fact many brain processes which remain in the dark or "behind the scenes" and whose activity is, accordingly, invisible. They

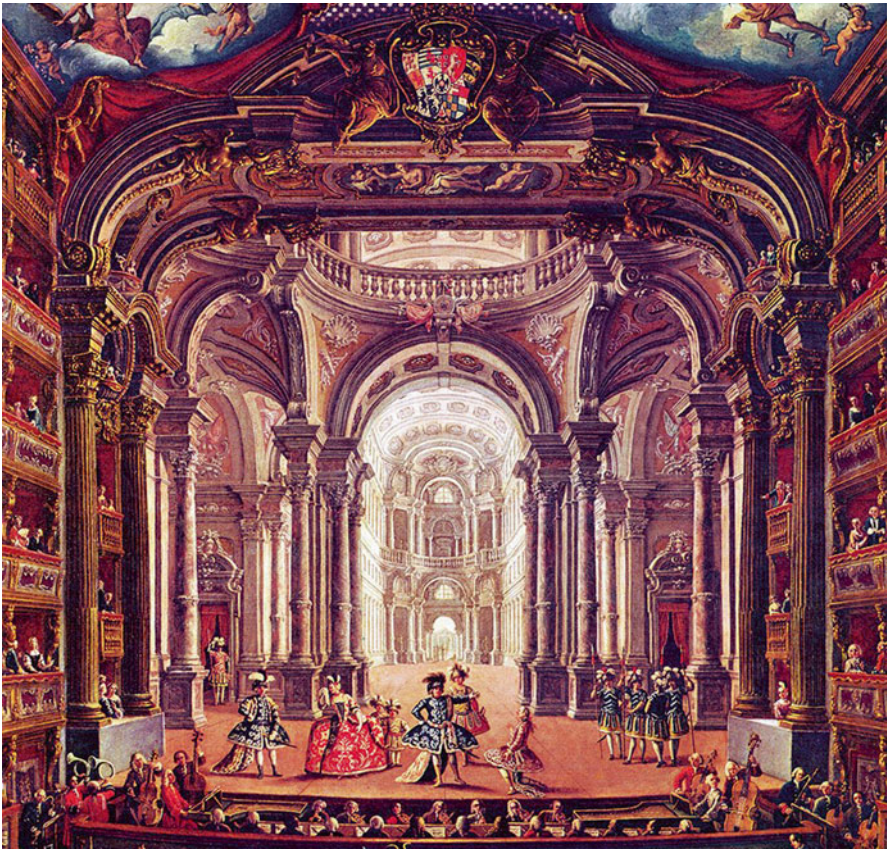


Fig. 16.1 Giovanni Michele Graneri (1708–1762), *The Teatro Regio in Turin*, oil on canvas (circa 1752), Palazzo Madama, Museo Civico d’Arte Antica, Turin, Italy (Image cropped by authors)

are however able to influence and orient consciousness. Baars calls these invisible brain processes *contexts*. Some examples are the executive functions, which operate like the theater director; the linguistic modules, which can be compared to script-writers (see Chap. 24 about the view that we are not conscious about the words we intend to express until they are said); and the dorsal cortical stream of the visual system, which shapes vision like stagehands shape the theatrical scenery. Despite this hidden activity, a thought or a content of experience must always enter the global workspace or, metaphorically speaking, be on the theater stage (Fig. 16.1) in order to be conscious.

It is important to highlight that the metaphor of the theater put forward by Baars is different from the concept of the *Cartesian theater*, in which a privileged observer (i.e., an “I”) within the brain watches the contents of the conscious experience on a sort of mental display (see Chap. 5 for a critique of this concept). In fact, within the global workspace theory, there is no dualistic assumption that “someone” is viewing the theater; likewise, there is no specific location or place in the

mind where experiences come to consciousness. The global workspace appears instead to be formed of different multiple networks which at times can or cannot be part of the global architecture. This theoretical construction shares some similarities with both Edelman's *dynamic core* (see Chap. 22) and Tononi's *dynamic complex* (see Chap. 28).

As we have seen, consciousness plays the fundamental role of "gateway" in the global workspace architecture, by enabling different cognitive networks to cooperate and compete in solving problems, such as the retrieval of specific memories, chunks of knowledge, etc. Baars calls this function of consciousness the *conscious access hypothesis*. He highlights that the property of allowing widespread access to brain resources differentiates consciousness from unconscious information processing, which by contrast is capable of limited elaboration, restricted to the brain sensory regions. Although some experimental paradigms have shown that the brain is able to process information unconsciously at an advanced level (e.g., the recognition of the meaning of words and the category to which objects belong), conscious elaboration of information remains fundamental to elicit spontaneous behaviors that the unconscious mind is not able to produce. Therefore, conscious processing appears to be more widespread and active, whereas unconscious processing appears to be more confined and passive.

In addition to this "gateway" function, albeit strictly related to it, Baars identifies other important roles for consciousness. The following are some functions among the most relevant ones identified by Baars.

Consciousness allows the comprehension of novel information. Consciousness seems to be needed to exploit specialized linguistic functions, such as syntax and semantics, in order to produce new combination of words.

Consciousness seems to be a necessary prerequisite of working memory. One needs to be conscious in order to report working memory elements like sensory inputs, rehearsal, and recall.

Consciousness makes possible many types of learning. To date, neuroscience has shown that long-term learning can occur only if the brain is conscious. Although memory mechanisms operate unconsciously, information must go through the focus of consciousness in order to enter episodic memory. Even implicit learning seems to require conscious information. In fact, implicit conscious paradigms ask participants to pay conscious attention to target stimuli. Moreover, consciousness is deeply involved in novel skills acquisition. During the process of learning a new skill, we must be conscious of every step of the new procedure, while after the skill has been learnt, automaticity tends to replace consciousness. Concomitantly, a large cortical activation is required during the learning procedure, while cortical activation is limited to a few areas when tasks are automatically performed.

Consciousness enables the voluntary control of actions. Consciousness seems to be necessary for implementing higher-order cognitive functions such as problem-solving and decision-making in order to voluntarily choose lines of conduct. Further, conscious goals can recruit motor systems in order to carry out actions.

Consciousness allows executive interpretations of behavior and, as a result, the emergence of a concept of self. Conscious input to frontal cortical areas might lead

to executive interpretation and control of behavior, which in turn might favor the construction of the sense of authorship (i.e., one's feeling of being the author of his/her own behavior).

American philosopher Ned Block (1996) is the originator of a famous distinction between two types of consciousness: *phenomenal consciousness* and *access consciousness*. The first type consists of subjective experiences and feelings; the second type consists of the global availability of information in a cognitive system so as to be used for reasoning, the rational control of action and speech. At first sight, it seems that Baars' *conscious access hypothesis* be very similar to Block's access consciousness. However, the two concepts should not be considered as equivalent. In Block's theoretical proposal, the phenomenal feature of consciousness is distinct and separate from the cognitive aspect of conscious processing, whereas in Baars' theory, it seems that the widespread access of information within the global workspace might lead both to the implementation of cognitive functions and to the phenomenal elaboration of sensory qualities, depending on the type of information which is "under the spotlight of the theater stage."

Since neuroimaging studies have shown how the conscious mind involves the activation of widespread neuronal networks, Baars' global workspace model has gained increasing attention within the neuroscientific community. The model has therefore been developed not only in its theoretical aspects but also and especially within a neuroanatomical context. In other chapters, we present in detail how three contemporary eminent neuroscientists – Stanislas Dehaene (see Chap. 19), Gerald Edelman (see Chap. 22), and Giulio Tononi (see Chap. 28) – have further developed the basic idea of the global workspace that consciousness needs extensive network activation in order to be produced and maintained.

Essential Bibliography

- Baars BJ (1988) A cognitive theory of consciousness. Cambridge University Press, Cambridge, MA.

In this book, the author puts forward a comprehensive and detailed account of the global workspace theory of mind. In the light of this model, Baars discusses a range of mental functions, such as language production, reasoning, problem-solving, decision-making, perception, and learning. Conscious and unconscious processes are compared and differentiated according to the degree of distribution of information into a vast assembly of neuronal networks. Unlike unconscious processing, the conscious mind appears to be strictly associated with the activation of the global workspace.

- Baars BJ (1997a) In the theater of consciousness: the workspace of the mind. Oxford University Press, Oxford

This book offers a brilliant introduction to the global workspace theory. Moreover, by combining theoretical psychology with neuroscience, the author is able to show how the metaphor of the theater can be a fruitful conceptual tool for understanding the nature of consciousness.

-
- Baars BJ (1997b) In the theater of consciousness: global workspace theory, a rigorous scientific theory of consciousness. *J Conscious Stud* 4:292–309
This article is a brief but detailed summary of the principal theoretical tenets of the global workspace theory of mind. The conscious access hypothesis is discussed and compared to the nature of unconscious processing, along with the metaphor of the theater and other important functions of consciousness.

A Neurobiological Framework for the Study of Consciousness

*The NCC is the minimal set of neuronal events
that gives rise to a specific aspect of a conscious percept.*
(A Framework for Consciousness)

Francis Crick was born on 8 June 1916 near Northampton, England, and died on 28 July 2004 in San Diego, California. He earned a degree in physics, and after World War II he began active in biology research. He became Nobel Laureate in 1962 together with James Watson and Maurice Wilkins for the discovery of the double helix structure of DNA. The second part of his outstanding career was devoted to the study of consciousness, as he held the post of J.W. Kieckhefer Distinguished Research Professor at the Salk Institute for Biological Studies in La Jolla, California.

Christof Koch was born on 13 November 1956 in Kansas City, Missouri, and earned a PhD in nonlinear information processing from the Max Planck Institute of Tübingen, Germany, in 1982. He subsequently spent 4 years as postdoctoral fellow at the Artificial Intelligence Laboratory and at the Brain Cognitive Sciences department at MIT. In 1986, he joined the California Institute of Technology (Caltech); in 2011, he became the Chief Scientific Officer of the Allen Institute for Brain Science.

The collaboration between Francis Crick and Christof Koch made important contributions to the study of consciousness for more than a decade, until the very end of Crick's life. Both scientists spent much of their research efforts to make the study of consciousness a respectable field of scientific inquiry. In the article *Towards a Neurobiological Theory of Consciousness*, published in 1990, they put forward a theoretical framework for the study of consciousness in order to show that this is a scientific tractable problem. According to Crick and Koch, the origin of consciousness can be found at the neural level and appears to have an intimate connection with other two brain properties: short-term memory and serial attention. Above all, the scientific study of consciousness requires two basic assumptions:

1. At any one moment, some active neuronal processes are associated with consciousness – the so-called neural correlates of consciousness (NCCs) – while others do not. Therefore, there must be a difference, which can be accounted for by neuroscience, between the activity of neurons which correlate with consciousness and those which do not.
2. All the different types of conscious experience – seeing a sunset, tasting wine, feeling pain, etc. – employ a basic common mechanism, so that if we can comprehend the mechanism involved in one conscious experience, then we will also be able to comprehend all other different modalities of conscious experience.

Based on these assumptions, a neurobiological framework for the scientific study of consciousness should be guided by the following hypotheses:

- (a) Everyone has a rough idea of what consciousness is, so that it is better to avoid a precise definition of consciousness because of the risk to give a premature definition.
- (b) It is also premature asking what the function of consciousness is until we can thoroughly explain its nature.
- (c) It is highly probable that some other animals, in particular higher mammals, could be conscious, so that experiments on these animals could be relevant for the study of consciousness. This implies that language is not an essential feature of consciousness, at least for its basic manifestations. It is, therefore, possible that consciousness correlates to some extent with the degree of complexity of the nervous system.
- (d) Self-consciousness is the self-referential aspect of consciousness and should be studied after having explained simpler forms of consciousness associated with sensory modalities.
- (e) A neural theory of consciousness should first attempt to construct a scaffold aiming to explain the most important aspects of consciousness and, then, attempt to propose more refined and inclusive models of conscious experiences.
- (f) The problem of qualia – that is, the subjective phenomenal aspect of consciousness (see Chaps. 1 and 5) – should be set aside until we achieve a precise understanding of the NCCs of a conscious experience, for example, seeing blue. However, a complete theory of consciousness should be able to account for how we can experience color altogether or how we can have any other sensation or feeling.

Thus, Crick and Koch's proposal is an attempt to build a theoretical structure of hypothetical assumptions in order to frame a possible scientific assault to the problem of consciousness. Within this framework, the problem can be formulated as follows: What is the special type of neuronal activity or process that correlates with consciousness? Is there a particular set of neurons involved in the generation of consciousness? And what is special about the connections and firings between these neurons?

In order to answer these questions, Crick and Koch first observe that structures in the midbrain or hindbrain, like the cerebellum, do not play any relevant role in the generation of the phenomenal consciousness and, thereby, can be excluded from the search for where these special neurons could be located. Likewise, the ascending reticular activating system (ARAS), which is responsible for regulating the level of alertness, can be considered as not essential for conscious experience. Undoubtedly, a degree of vigilance must be present in order to have consciousness, but to equate consciousness to the ARAS activity would be like believing that the programs on television depend on the TV electrical supply. According to Crick and Koch, it is highly plausible that the neocortex and other brain structures intimately associated with it (including the thalamus, basal ganglia, and claustrum) can form a large network of neurons whose activity is of fundamental importance for any aspect of consciousness.

Among the different sensory modalities, Crick and Koch choose to focus on vision. Neuroscientific research on the mammalian visual system revealed that different brain areas are involved in the construction of a visual image. In general, different cortical regions respond to different visual inputs. For example, neurons in the MT area respond mainly to motion and depth while those in area V4 to color and shape. Given these facts, it is important to understand how the brain is able to generate a coherent single visual picture (the so-called binding problem). In other words, it remains to be explained how we can perceive the world not through different and separate sensations, but through a unitary conscious scene made up of various sensory inputs, feelings and emotions, needs an explanation. Thus, the fundamental issue regarding consciousness appears to be how the brain is able to bind together in a consistent fashion the different streams of information elaborated by neuronal networks which respond to different aspects of the perceived objects.

Crick and Koch's hypothesis is that, at any one time, large coalitions of neurons in different cortical areas cooperate in order to sustain a common global activity, which finally corresponds to visual awareness. Both the processes of attention and working memory are also essential for the enhancement and maintenance of this global activity. Scientific research on the cortical visual areas showed that neuronal activity within these regions appears to be synchronized at approximately 40 Hz. Based on this, Crick and Koch propose that neurons can bind representations together when firing at about 40 Hz. In a sense, these common neuronal oscillations could be the hallmark of consciousness, which would emerge whenever a set of neuronal coalitions synchronously play this score. In other words, the coalitions of neurons bind together when they fire the harmony of consciousness, just as the legendary figure of Erin is chained to the rock emerging from the sea while playing her harp (Fig. 17.1).

In their 1990 article, Crick and Koch suggested that these rhythmic oscillations do not encode additional information but join previously elaborated information into a coherent percept. However, in a subsequent paper published in 2003 – *A Framework for Consciousness* – they refined their neurobiological scaffold and declared that they no longer believed that the synchronized neuronal firings at 40 Hz is a sufficient condition for the NCC. They proposed that this sort of

Fig. 17.1 Thomas Buchanan Read (1822–1872), *The Harp of Erin* (1867), oil on canvas, Cincinnati Art Museum, Cincinnati, Ohio, US (Image cropped by authors)



synchronization could play a role in the emergence of consciousness only in case there are multiple rival coalitions of neurons which try to become predominant over each other. In this view, a synchronized coalition would increase its effectiveness and win the battle for accessing consciousness. Thus, synchronization at 40 Hz would not provide a convincing solution to the “binding problem,” but in a simpler way the binding would occur when small sets of neurons become all members in a particular coalition. The different neurons within a coalition support each other and sustain a global process by increasing the activity of their fellow members, so that the coalition which has the more intense and prolonged activity can prevail over the others and embody the contents of consciousness.

Coalition of neurons can be either conscious or unconscious. The latter ones are called by Crick and Koch *zombie modes*. The function of consciousness would allow the individual to think and plan more complex behaviors, while the zombie modes would respond to environmental inputs in a rapid and stereotyped manner, like cortical reflexes. Moreover, Crick and Koch introduce two central concepts for the generation of consciousness, in particular with regard to visual awareness: the *snapshot* and *threshold*. A conscious visual scene or percept would emerge through a succession of static snapshots that can be maintained for a certain time above a

threshold of neuronal activity characterized by a constant firing rate. According to this view, a snapshot would be formed by a specific coalition of neurons.

Another property of brain networks, which is called *penumbra* by Crick and Koch, is strictly associated with these concepts. The penumbra is the neuronal activity (including synaptic effects and firings rates) which is not part of an NCC, but it is nonetheless influenced by it. Thus, the penumbra can include past connections or associations of NCCs, such as the expected consequences of these NCCs, possible plans of movements, memories, and so forth, which can all be indirectly activated by the NCCs. For instance, the taste of a madeleine cake represented in a specific NCC would induce the past memory of the moment when this taste was perceived for the first time. As the NCC shifts, the penumbra can too become conscious. In fact, coalitions of neurons that form NCCs are intrinsically dynamic and constantly changing. Snapshots are relentlessly constructed by the brain and follow one another in such a rapid manner that we have the illusion of a continuous conscious flow.

During the last decade, Koch gradually acquired a broader perspective on consciousness. Specifically, he endorsed Tononi's theory of integrated information (see Chap. 28) and developed his speculation with intriguing metaphysical suggestions, with leanings to panpsychism, the philosophical doctrine which attributes a modicum of consciousness to all matter. He still remains a reductionist neuroscientist but, as he has written in his last book, a romantic one. Thus, the difference between him and Crick could be described as follows. If for Crick we are nothing but what our neurons do, for Koch we are more similar to a very special process of elaborated information, which undoubtedly occurs in the brain but could theoretically occur in other types of functional organization.

Essential Bibliography

- Crick F (1994) *The astonishing hypothesis: the scientific search for the soul*. Charles Scribner's Sons, New York
This book is a pioneering and thought-provoking attempt to give a reductionist account of consciousness. The astonishing hypothesis revealed by Crick is that what we are, from our joys to our sorrows, from our memories to our hopes, is no more than the products of neuronal activity in the brain.
- Koch C (2002) *The quest for consciousness*. Roberts & Company Publishers, Englewood, Colorado
This book presents one of the most lucid syntheses of the neuroscientific approach directed to identify the neural correlates of consciousness. The state of the art in the neuroscientific research is described by the author using a clear and fascinating style.
- Koch C (2012) *Consciousness: confessions of a romantic reductionist*. MIT Press, Cambridge, MA.
While the first book of Christof Koch adopted a scientific and objective perspective, this second one shows a more intimate flavor. Following the common thread of the neuroscientific discussion, Koch leads the reader through his memories and intriguing philosophical ideas.

Consciousness, Emotions, and Self

I believe the contents of consciousness that we can access are assembled mostly in the image space of early cortical regions and upper brainstem, the brain's composite "performance space". What happens in that space, however, is continuously engineered by interactions with the dispositional space that spontaneously organizes images as a function of ongoing perception and past memories. At any given moment, the conscious brain works globally, but it does so in an anatomically differentiated manner.
(Self Comes to Mind)

Antonio Damasio (born on 25 February 1944 in Lisbon, Portugal) is a leading neuroscientist who has fundamentally contributed to the field of consciousness studies. Damasio studied medicine and neurology at the University of Lisbon Medical School, where he also completed his doctorate. He subsequently moved to the United States and conducted research in the fields of behavioral neurology and neuropsychology at the Aphasia Research Center in Boston, under the supervision of American neurologist Norman Geschwind. His research interests then focused on the physiology of emotions, with regard to both their nature and function as well as their subserving neuronal substrates. The results of his studies led him to formulate the *somatic marker hypothesis*, an original theory about how emotions can affect the process of decision-making and social cognition. Damasio was M.W. Van Allen Professor and Head of Neurology at the University of Iowa Hospitals and Clinics. Since 2005, he has been David Dornsife Professor of Neuroscience at the University of Southern California, where he currently heads the Brain and Creativity Institute. He is also Adjunct Professor at the Salk Institute, La Jolla, California.

Damasio has repeatedly highlighted that the brain and its functions, including consciousness, cannot be studied in isolation. In other words, each brain activity and cognitive function should always be regarded as part of a complex interplay between the organism and its environment. In light of this, mental states appear to be deeply embodied and cannot be accounted for without considering the reciprocal influences between the body and its surroundings. According to Damasio, unrecognizing that mind, body, and environment reciprocally shape each other was Descartes' worst error (see Chap. 6 for a discussion about Descartes' theoretical position).

According to Damasio, every interaction between the brain and its internal and external milieu results in changes in the way the mind represents the world. These changes are codified by different kinds of dynamic maps. Damasio identifies three varieties of maps: *interoceptive maps*, *proprioceptive maps*, and *exteroceptive maps*. The interoceptive maps are representations of the organism's internal structure and state. These representations denote the functional condition of body tissues, such as the degree of contraction or distension of musculature. The proprioceptive maps are representations of specific body components, such as joints, viscera, and striated muscles. Finally, the exteroceptive maps are representations of states of affairs in the external world. These representations signify events or objects that stimulate the sensory systems, such as the retina, the cochlea, and the mechanoreceptors of the skin.

Damasio calls *images* all the sensory mapping which can become the subject of conscious experience. Thus, images can come from any sensory modality, not only vision. Within this category, he identifies *primordial feelings*, which are the images related to the organism's internal state, and *body feelings*, which are the images resulting in the combination of primordial feelings with maps related to the other aspects of the organism. Thus, according to Damasio, feelings are a special kind of representation, in virtue of their unique relation to the body. In his terms, feelings are naturally *felt* images. This construction introduces a clear-cut distinction between *emotions* and *feelings*. Emotions are simple behavioral reactions, largely associated with automated programs of actions. These actions are carried out in our bodies and include facial expressions, postures, and changes in the viscera. On the other hand, feelings are composite perceptions of what is going on into the body when emotions occur. Thus, feelings are images of actions rather than actions per se. In other words, emotions are reactions associated with ideas and certain modes of thinking, while feelings are perceptions of what the body does during emoting, along with perceptions of the related state of mind. The process of interoception, therefore, characterizes the exclusive relationship between brain and body.

In this view, consciousness appears to be a property of the brain that emerges when the interplay between brain, body, and environment reaches a certain degree of sophistication. Damasio identifies three main ingredients for a full-blown state of consciousness: *wakefulness*, an *operational mind* and a *sense of self*. First, a state of alertness or wakefulness is a fundamental prerequisite for being conscious. Still, consciousness and wakefulness are distinct brain processes, as it is shown in the neurological condition known as vegetative state, in which patients clearly present alternating patterns of sleep and wakefulness, but no signs of consciousness. Second, consciousness needs an operational mind, that is, a mind that is able, as we have seen, to work as effective interface between the body and the world. Third, the mind should also be able to produce a sense of self as the protagonist of the experience. Damasio summarizes the entanglement of these three components by saying that the distinctive feature of consciousness is "the very *felt* thought of you."

Thus, in Damasio's account of the mind, consciousness results from the complex interaction of different cognitive and emotional processes. These brain processes can produce a composite synthesis of different images, which in turn represent the objects or contents of consciousness as well as the sense of self. In order to construct a self, Damasio identifies four basic components: *perspective*, *ownership*, *agency*, and *primordial feelings*. First, the self needs a perspective or a standpoint from which the objects of the mind are mapped and represented in images. Second, the self is built on the sense of ownership, the feeling that the objects represented in the mind belong to the self. Third, the self is strictly intertwined with the sense of agency, the feeling that the actions carried out by the body are controlled by the body's mind. Fourth, the self must be enrooted in an internal world formed of bodily feelings.

The aggregation of these four elements makes a self in its simplest version, which is called by Damasio *protoself*. The protoself is an interoceptive integration of neural patterns that constitute a map of the current organism's physical condition. The internal maps are also integrated with the external sensory inputs. Integrations occur both at the level of the brainstem (nucleus of the solitary tract, parabrachial nucleus, periaqueductal gray, area postrema, hypothalamus, and superior colliculus) and at the level of the cerebral cortex (insular cortex, anterior cingulate cortex, frontal eye fields, and somatosensory cortices). In turn, the protoself is the basis on which a higher-order self (*core self* in Damasio's terminology) can be constructed. A core self is produced when the primordial feelings are transformed into feelings of knowing that the protoself is engaged in changes caused by the perceived objects of the external world. Each time the organism encounters an object, its protoself is changed by this encounter. In turn, changes in the protoself lead to the emergence of the core self, which is the protagonist of a narrative in which the organism connects with the events that it is involved in. The development of the core self, thus, requires the bonding of the modified protoself to the object that caused the modification. This process occurs in a simple cycle: an object sensorily engages the body from a specific perspective, this engagement in turn causes the body to change, and, finally, the presence of the object is felt, recognized, and made salient. According to Damasio, the narrative story of such continuous engagements is the way the self comes to mind. This story develops into a succession of images, resulting in the modifications of the protoself, as well as the emotional responses or feelings stirred by the perceived objects.

At the highest level of this hierarchy of selves, Damasio puts what he calls the *autobiographical self*. The autobiographical self is literally our autobiography made conscious. It is formed of personal memories, life experiences, and future plans – no matter how specific or vague. It is what constitutes the biographical identity of a person and, similarly to the protoself and the core self, is a highly dynamic entity which undergoes constant reconstruction. It is as if the autobiographical self were the subject of a gallery of self-portraits that our minds depict along the years, similarly to the famous self-portraits painted by Dutch painter Rembrandt along the years (Fig. 18.1).



Fig. 18.1 Rembrandt Harmenszoon van Rijn (1606–1669), three among his several self-portraits. (a) Self-portrait of a young Rembrandt laughing (circa 1628), Getty Center, Los Angeles, California. (b) Self-portrait of a middle-aged Rembrandt (circa 1640), Norton Simon Museum of Art, Pasadena, California. (c) Self-portrait of an old Rembrandt with two circles (between 1665 and 1669), Kenwood House, London. All the three images have been cropped by authors

According to Damasio, the mechanism through which the autobiographical self is generated is extremely complex. Roughly speaking, three steps are needed: (1) a set of biographical memories, which are represented as images; (2) an interaction between these biographical images and the protosef, which is in turn modified by

this interaction; and (3) a transiently coherent pattern of states of the core self, which are generated in a pulse-like fashion, as a result of the interaction between the biographical images and the protoself. The brain structures involved in generating the autobiographical self are the brainstem, the thalamus, and the cerebral cortex.

In Damasio's view, consciousness develops in parallel to the sense of self. Thus, alongside the construction of the core self and the autobiographical self, Damasio identifies two types of conscious elaboration: *core consciousness* and *extended consciousness*, respectively. The first one is strictly intertwined with the core self, while the second one corresponds to the autobiographical self. Core consciousness is related to the sense of self at a specific moment, its scope is restricted to the here and now. It does not deal with the future but only with instant glimpses of the present. As the core self, core consciousness arises by virtue of the interaction between an object and the body. When the interaction occurs, the brain creates a chain of different images: an image of the body's internal state, an image of the object, and an image of the modification of the internal state caused by the encounter with the object. The brain also produces a higher-order image which includes the other ones and enters the space of consciousness as the feeling of an experiencing self. Extended consciousness, by contrast, is related to the sense of the autobiographical self as well as to the individual's identity. It is extended in time and deals equally with the past, the present, and the future. It cannot arise in the absence of core consciousness and working memory, because these two processes are required in order to form a sense of self in the very act of knowing, which is, as we have seen, what Damasio calls the autobiographical self. Extended consciousness can also be enhanced by language and long-term memory.

Damasio has the merit of putting forward one of the most interesting and sophisticated theories of consciousness based on neuroscientific grounds. Two points in his theoretical position are particularly worth discussing. The first point is the claim that some brain regions are more important than others in the generation of consciousness. Although this view is commonly accepted within the community of neuroscientists, Damasio intriguingly suggests the essential role of the brainstem in constructing the protoself, which therefore might be the basis upon which the conscious mind can be created. The second point is the idea that consciousness results in a progression of sensitivity, so that it can yield ever more sophisticated pictures of the world. Accordingly, consciousness and emotions appear to be so entangled that without feelings there could be no consciousness at all.

With regard to the first point, there seems to be an order of precedence between self and consciousness, with the self coming first, at least in its simplest form (protoself). Still, it could be argued that neuropsychological evidence suggests the possibility of conscious perceptions which do not appear to be accompanied by any sense of agency or ownership. With regard to the second point, feelings and emotions are essential ingredients for the recipe of consciousness. Arguably, consciousness might need a mechanism of feedforward and recurrent networks capable of generating states which could not necessarily have the forms of emotions. Only future research

will be able to tell whether these two important theoretical standpoints are really destined to play a key role in the science of consciousness.

Essential Bibliography

- Damasio A (1994) *Descartes' error: emotion, reason, and the human brain*. Putnam, New York.
This intriguing book explains how the philosophical position of Cartesian dualism can be appraised in light of modern neuroscience. The author claims that Descartes' support to the concept that mind and body are to be thought of as distinct and separate substances was a deep error, which had deleterious consequences for the scientific study of the mind, including the idea of separating rationality and emotions.
- Damasio A (1999) *The feeling of what happens: body and emotion in the making of consciousness*. Harcourt, Orlando.
The author leads the reader through a fascinating investigation on how the brain can create a conscious picture of what happen both in the inner and external worlds. The distinction between feelings and emotions is clearly discussed, alongside the different types of conscious elaboration (i.e., core and extended consciousnesses).
- Damasio A (2010) *Self comes to mind: constructing the conscious brain*. Pantheon Books, New York.
This book explores the processes by which the brain is able to construct a permanent sense of self and to link it to conscious experiences. Different stages in the generation of the self are examined in detailed with attention to a wide range of aspects, from the theoretical to the neuropsychological and neuroanatomical ones.

Consciousness Is Global Information Sharing

...my own theory of a neuronal workspace [...] proposes that a conscious state is encoded by the stable activation, for a few tenths of a second, of a subset of active workspace neurons. These neurons are distributed in many brain areas, and they all code for different facets of the same mental representation.
(Consciousness and the Brain)

Stanislas Dehaene (born on 12 May 1965) is a French leading neuroscientist whose work has primarily focused on the neuronal bases of reading, numerical cognition, and consciousness. Dehaene is trained as a mathematician and in 1985 gained his Master's degree in Applied Mathematics and Computer Science at the University of Paris VI. His interests then shifted to neuroscience and cognitive psychology and in 1989 received his PhD in Experimental Psychology from the École des Hautes Études en Sciences Sociales (EHESS) in Paris. He began to collaborate with prominent French neurobiologist Jean-Pierre Changeux on computational neuronal models of human cognition, including working memory and task control. Dehaene conducted his research in these fields at the Cognitive Sciences and Psycholinguistics Laboratory of the Institut National de la Santé et de la Recherche Médicale (INSERM – Institute of Health and Medical Research). From 1992 to 1994, he spent 2 years as postdoctoral fellow at the Institute of Cognitive and Decisions Sciences, University of Oregon, under the supervision of American psychologist Michael Posner. He then returned to France, where he developed an active research group. In 2005, he was appointed as Chair of Experimental Psychology at the Collège de France in Paris. He is also the current Director of INSERM Unit 562 “Cognitive Neuroimaging.”

Dehaene's theoretical approach to consciousness has been inspired by Bernard Baars' workspace theory (see Chap. 16) and can be considered as one of the most promising refinements of this model within a neuroscientific framework. The basic tenet of this approach – which is called by Dehaene *neuronal global workspace* – is that consciousness makes information globally available within the brain. The global workspace is an evolved architecture for extracting relevant information from the environment and for spreading it to higher-order brain decision systems as well as language processors in order to name it. Thus, as soon as information enters this specific broadcasting circuitry, it becomes conscious. In other words, according to Dehaene, consciousness is simply brain-wide information sharing. In light of

that, the hallmark or the most characteristic feature of consciousness is its flexible dissemination of information.

The main elements which form the global workspace are different coalitions of neurons or cell assemblies, associative brain areas, and convergence zones with reentrant circuits. These cell assemblies often compete with each other in order to gain a temporary dominance within the global workspace, so that, when a subset of workspace neurons are stably activated, a conscious state is encoded. These neurons, distributed in many brain regions, codify for different features of the same mental representation. Each cell conveys a little bit of information about a specific scene, but, collectively, all the cell assemblies are able to represent an infinite repertoire of images and thoughts. Out of this immense potential set, a single mental pattern is selected and becomes the focus of attention and consciousness. In this very moment, all the neuronal coalitions which codify for that mental pattern activate in partial synchrony under the control of a subset of prefrontal cortex neurons.

Dehaene subtly highlights that the cell assemblies which remain silent or less activated also encode pieces of information. Their silence implicitly reveals to the other coalitions that the features they fire for are not present or are irrelevant to the current mental representation. Thus, a conscious content is equally defined by its active neuronal groups and its silent ones. As Dehaene points out, the shaping of an idea can therefore be compared to the sculpting of a statue. Italian Renaissance artist Michelangelo wrote in a famous sonnet that the sculptor does not add elements as painters do on canvases in order to draw pictures but does the exact opposite. The sculpture is made *in levare* (by chipping away) most of the marble block so as to progressively expose the artist's vision (Fig. 19.1). Similarly, our brains sculpt a mental scene by silencing most of the cell assemblies within the neuronal global workspace and by keeping active only a small fraction of them. These active neuronal groups literally shape our conscious thoughts.

On the basis of the global workspace theory and the empirical evidence gathered thus far by neuroscience, Dehaene identifies four main signatures of consciousness. These signatures are four characteristic features of the functioning brain, which can be considered as reliable markers that something has been consciously elaborated. The first signature is an amplification of sensory brain activity, which progressively acquires strength so as to spread across multiple regions of the cerebral prefrontal and parietal lobes. The second signature is a characteristic EEG pattern: the P3 wave (so called because it is the third positive peak after the presentation of the stimulus) or P300 wave (so called because it appears around 300 milliseconds after the presentation of the stimulus). During experimental EEG monitoring, this pattern has been frequently associated with having conscious access to specific information. The P3 wave is only the beginning of an avalanche of activity that suddenly bursts into a wide range pattern of high-frequency oscillations. This extensive activity, which Dehaene calls *global ignition*, is the third signature of consciousness. When the firing rates of neurons exceed a certain threshold, the activity becomes self-reinforcing, so that some neurons excite others which, in turn, reciprocate the excitation. The result is a self-sustaining state of reverberating cell assemblies. This leads to what Dehaene calls the *brain web*, which is the fourth signature of

Fig. 19.1 Michelangelo Buonarroti (1475–1564), *Young Slave* (1520–1523), marble, Accademia delle Belle Arti di Firenze, Florence, Italy (Authors' picture)



consciousness. The brain web is formed by massive exchanges of reciprocal signals (codes of information) across different and distant regions within the brain. This information processing typically occurs by virtue of highly synchronized brain signals fluctuating along bidirectional connections.

Reentrant connections or neuronal loops are of extreme importance in the architecture of the brain web. Their extension determines how global the neuronal workspace that creates our inner conscious world is. Although it is possible that in principle any kind of neuronal loop, no matter how little, can engender a shred of awareness, Dehaene claims that only the long-distance loops, running across the prefrontal and parietal regions, can bring about a full-blown conscious experience. In his view, the reverberating activity per se, albeit necessary, is not sufficient for consciousness. In case of consciousness, information must also be widely

distributed throughout the brain along extended routes. On the contrary, short neuronal loops would suffice for unconscious processing.

According to the global neuronal workspace, consciousness emerges when a coherent brain web ignites in a self-sustaining way. Conscious processing seems therefore to play a specific function in the brain computational economy. Dehaene suggests that its main functional role might be to select and bring to the subject's attention one among the innumerable possible interpretations of the world. In light of that, consciousness could be considered as the brain's discrete measurement device, a device able to collapse the large number of unconscious data into a coherent conscious scene. Whereas unconscious processing is quick and operates in a massively parallel fashion, conscious elaboration is by contrast slow and serial. At any given time, only one sample of the world enters the focus of our attention. This allows us to keep in mind lasting ideas, which in turn are the elements that we use to make decisions. Conscious access to information seems therefore to be essential for reasoning, because it makes possible to follow rational strategies of thought. However, it could be argued that consciousness does not seem necessary for following rational strategies of behavior, as Dehaene claims. For instance, some computer programs designed for playing chess can perfectly follow a rational game strategy so as to play at the same level of the great masters.

Dehaene's position assigns consciousness a causal role which is crucial in an evolutionary perspective. Without consciousness, we could not simply do the things we do, and we could not be the human beings we are. Consciousness is, in other words, fundamental in order to produce our volitional behavior. This approach is in deep contrast to the position that considers the conscious mind as an epiphenomenal by-product of the brain. The theory of epiphenomenalism has gained supporters in both philosophy of mind (see Chap. 8 for a critical appraisal of this doctrine) and neuroscience (see Chap. 22). Conversely, Dehaene's view openly challenges any epiphenomenalist account of consciousness. This is because he puts forward a computational theory of consciousness. Accordingly, if consciousness can be reduced to computational operations within the brain, it necessarily follows that consciousness too is a causal process.

Another important consequence of Dehaene's approach is the rebuttal of the concept of phenomenal consciousness. According to him, thinking about conscious experience as having a phenomenal feature is misleading, confusing, and, what is worse, conducive to the slippery slope that leads to dualism. However, it could be argued that the computational and cognitive sides alone cannot account for the whole nature of conscious phenomena. Many authors presented in this book argue that consciousness does not seem to be completely reducible to a specific set of computational operations within the brain.

Philosophical arguments suggest that consciousness might have a facet that escapes descriptions expressed in computational terms only. In fact, computational theories fall short when they try to account for an aspect which appears to be deeply connatural to any conscious state, namely, the qualitative flavor which constitutes the particular perspective that each conscious creature has on the world. In other

words, consciousness appears to be, first of all, an intimate contact with the world. And as we have seen in the previous chapter, this contact might heavily rely on emotions; and insofar as emotions permeate our thoughts, they too are essential ingredients in the process of making decisions.

Despite being grounded in neurobiological substrates, Dehaene's theory reveals a strict computational approach. Accordingly, Dehaene keeps an open mind on the possibility that artificial consciousness could in principle be achieved, provided that future machines have three critical functions: (1) flexible communication, (2) plasticity, and (3) autonomy (but see Chap. 15 for a different opinion on the theoretical possibility of an artificial conscious mind). With regard to communication, at present most computer programs have no chance of exchanging their respective knowledge. However, if the global workspace theory is correct, information is to be widely distributed and exchanged in order to be conscious. With regard to plasticity, each program should be able to shape itself in a brain-like fashion according to a powerful learning algorithm. With regard to autonomy, an artificial global workspace should constantly produce a spontaneous or autonomous activity, a never-ending flow of flickering internal states.

Finally, in Dehaene's theory, the concept of self does not have a privileged status with respect to consciousness. Therefore, it would appear that someone does not need to be self-aware in order to be conscious of something. Self-consciousness would be just another form of conscious experience, similar to all the others, with the only difference that in self-consciousness the content of the experience is not a specific percept (such as a color or a sound) but an aspect of oneself (such as a bodily state). In this case, consciousness does not focus on the information coming from the sensory routes but on one among the various mental representations of the "I" concerning the body, behavior, feelings, and thoughts. This view contrasts with Damasio's claim that a concept of self is crucial in order to develop consciousness (see Chap. 18). At this point in time, both possibilities are equally arguable, and only advances in neuroscientific research will be able to tell which of the two hypotheses provides the best model to understand the relationship between the self and consciousness.

Essential Bibliography

- Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace network. *Cognition* 79:1–37.
This article, coauthored with French neurologist Lionel Naccache, provides an excellent technical discussion of the neuronal global space theory. The problem of the nature of consciousness is taken into consideration from different perspectives, including the philosophical and the empirical standpoints. Special attention is paid to the theoretical bases of cognitive neuroscience of attention and consciousness.

- Dehaene S (2014) *Consciousness and the brain: deciphering how the brain codes our thoughts*. Viking, New York.
This book is a fascinating exploration of the state of the art of neuroscientific research about consciousness, attention, and unconscious processing. Dehaene lucidly argues that consciousness is a problem which can be tackled using a scientific methodology. Four signatures of consciousness are proposed and combined with the theory of the global neuronal workspace.

The Evolution of Human Consciousness

The nature and range of human conscious experience are no longer a biological given. Rather they depend on a somewhat unpredictable chemistry of brain and culture, whereby the processes of mind can be endlessly rewritten and rearranged by cultural forces.

(A Mind So Rare)

Merlin Donald (born on 17 November 1939) is a Canadian psychologist and cognitive neuroscientist. In 1968, he gained a PhD in Neuropsychology from McGill University; in 1972, he joined the Faculty of Queen's University at Kingston, Ontario, where he was Professor of Psychology, now Emeritus. In 2005, he contributed to found the Cognitive Science Department at Case Western University and was the chair of that Department until April 2013. He is currently Adjunct Professor at Case Western University.

Merlin Donald is well known as being the author of two books on human cognition and consciousness: *Origin of the Modern Mind* and *A Mind So Rare*. His theory about the origin of the human mind and its development lies firmly within the framework of evolutionary theory. His hypothesis is that the modern human mind evolved from the primate mind through progressively more sophisticated adaptations, each of which brought about a new representational system. As each representational system has remained intact within the human mental organization, now the modern mind appears to be a mosaic of different cognitive structures, all derived from earlier stages of human development but preserved as evolutionary gains.

Donald's hypothesis is that the modern human mind developed throughout three major stages or cognitive revolutions: mimetic, mythic, and theoretic. Each of these transformations marked a specific period and culture. Thus, according to Donald, the development of the human mind should not be considered as a continuous or unitary process but rather as a nonlinear progression characterized by radical evolutionary changes. The first and the second transitions derived from biological changes, specifically a rapid increase in the cerebral volume and the descent of the larynx, which made possible the emergence of spoken language as we know it. The third transition, however, was not dependent on biological changes but was characterized by cognitive enhancement resulting from the support of new technological tools. It basically started in the later Upper Paleolithic, with the creation of the first permanent visual symbols, and is still under way.

Each transition introduced particular cognitive instruments to the human mind. The mimetic transition introduced the capacity of mimesis, which made possible to form representations that could be voluntarily retrievable. This mimetic ability allowed hominids to use their whole body as a representational device and had the important property of being autocued or self-triggered in rehearsal loops. In other words, hominids could voluntarily make use of mimetic actions in order to represent reality and then purposely retrieve these actions with their mimetic skills. Mimesis is in fact dependent upon a memory system that can recall, repeat, and refine sequences of movements, which are in turn guided by perceptual models of the body located in its surrounding environment. Donald suggests that the retrievable body memories were the first true mental representations in the history of mankind, as well as the basic form of reflection formulated by our ancestors. What is more, the mimetic capacity set the conditions for the later emergence of language. In fact, motor mimesis would have allowed a quasi-symbolic communication and the creation of a simple shared semantic milieu.

The mythic transition was characterized by the capacity of lexical invention. Human imagination literally flourished and created new symbolic and phonological combinations of words. This rapidly led to the evolution of a language system, with the development of complex metalinguistic skills that set the rules for the use of words. The collective product of this linguistic thrust was storytelling or narrative thought. Language was used to construct conceptual models of nature. During the mythic stage, the human mind was bound to concrete facts, whereas now it was able to reach abstract reasoning by using metaphors and thereby derive general principles from natural phenomena. The myth therefore became the prototypal and fundamental mind tool, by which knowledge over natural and social phenomena can be conveyed and spread. The most important and famous myths of the ancient past are currently studied in schools of the Western world, such as the tales narrated by the ancient oral tradition identified in Homer in the *Iliad* and the *Odyssey*. Figure 20.1 illustrates one of these ancient myths, the abduction of Helen, which was the cause for waging war against Troy.

The theoretic transition was triggered by the invention of external memory devices, records, and graphic symbols, which developed in complex writing systems. This marked the passage from oral culture to external theoretic culture. Moreover, the use of external storage systems required a redeployment of brain resources. The human brain underwent a reorganization in which literacy-related cerebral modules were established. Donald suggests the term *exogram* (i.e., recorded information on external supports) to complement the notion of *engram* (i.e., recorded information on biological supports). Writing systems made possible the construction of complex theories about virtually everything, hugely increasing the human potential for abstract reasoning.

Donald points out that the essential cognitive adaptation underlying each of the three great cognitive transitions in human evolution was a new system of memory representation. Moreover, during each transition, the previous cognitive instrument was encapsulated by the new representational structure. As a result, consciousness can take many forms within the context of mental architecture proposed



Fig. 20.1 Maerten van Heemskerck (1498–1574), *Panorama with the Abduction of Helen Amidst the Wonders of the Ancient World*, oil on canvas (1535), The Walters Art Museum, Baltimore, Maryland (Image cropped by authors)

by Donald. In fact, consciousness depends upon the momentary locus of control associated with a specific representational system. Within an episodic system, consciousness is situation bound and concrete. There is no possibility of long-term conscious planning, and the conscious focus is almost totally absorbed by the present moment. Within a mimetic system, consciousness starts to have a social dimension and regulate mimetic skills: gestures can be mimed or imitated, and the capacity of producing metaphors enriches communication, even though the cognitive sphere is still dominated by factualness. Within a mythic system, instead, consciousness extends beyond the situation-bound limit and incorporates the autobiographical self. Conscious reflection is permeated by narrative thought and can encompass representations of the past, present, and future. Finally, within a theoretic system, contents of consciousness are massively expanded into external symbolic networks.

In addition to these three functional aspects of consciousness, Donald identifies three basic levels of awareness, which appear to be correlated with the anatomical and functional development of three brain regions: the primary, secondary, and tertiary cortices. The primary cortex includes sensory areas and the primary motor cortex, both directly connected to the peripheral nervous system. The sensory areas elaborate signals coming from the senses and construct the basic features of sensations. The primary motor cortex sends signals to muscles in order to perform actions. The first level of awareness is associated with the activity of this primary neuronal network and is characterized by the selective binding. Living creatures that reached this level of consciousness are in fact able to simultaneously combine different perceptual aspects of the external objects in coherent and unified

percepts. In turn, this basic property of consciousness, which is guided by attention, is at the root of more abstract forms of conscious perceptions regarding events extended in time.

The secondary cortex, which receives inputs from the primary sensory areas, includes visual, auditory, and somatic regions and can perform advanced perceptual functions, such as face recognition, objects discrimination, and resolution of sound patterns. It also includes the premotor cortex, which is involved in the execution of motor programs and action planning. The second level of awareness appears to be associated with the activity of these cerebral areas in conjunction with the brain system that supports short-term memory. Both the ability to bind different perceptual features and the temporal dimension of the perceptual scenery, which can reside within memory over several seconds, are broadened.

The tertiary cortex has no connections with the peripheral nerves or the primary cortex. Instead, it has reentrant connections with the secondary cortex, other tertiary areas, and important subcortical regions, such as the limbic system. Tertiary regions are completely devoted to abstract cognitive processing and can thereby control and supervise all the sensory and motor elaboration, as well as perform executive functions, such as metacognition and complex planning. The activity of these brain regions develops what Donald calls the third level of awareness, which is related to the intermediate- and long-term control and regulation of both thought and behavior. This level of conscious capacity further expands the range and variety of experience. With the help of imagination, the perception of space extends far beyond the immediate perceptual horizon, and time is perceived as a continuous stream of events. Finally, this type of awareness makes possible self-reflection, that is, the idea of oneself acting in a three-dimensional world. This sense of self is constructed and maintained by interweaving autobiographical events which are stored in the brain network supporting long-term memory.

Donald describes a complex vision of the origin and nature of consciousness, under which he clusters different brain functions and aspects of human behavior. His central tenet is that consciousness cannot be singled out and understood in isolation but rather is to be considered as the dynamic product of both biological and cultural forces that are inextricably intertwined with each other. According to Donald, the evolution of different modes of representing reality has paralleled and shaped the evolution of consciousness. Each stage in human cognition was developed by expanding specific sets of cognitive skills and reprogramming the mind accordingly. One of the most important aspects pointed out by Donald is that certain thoughts could not be conceived without language. In other words, some thoughts can enter the mind's conscious space only if they are incorporated in symbols. The potential of consciousness reaches its peak through symbolic communication.

Another important point stressed by Donald is that the human conscious mind has gradually acquired across time an ever more refined integrative function. This could constitute a hint for a biological reason for consciousness. New skills and information are never assembled unconsciously but rather require entering memory consciously. The brain seems to need a conscious supervision in order to assemble

and learn new chunks of information. Moreover, this conscious supervision appears to be necessary to control the hierarchies of brain modules that implement specific cognitive skills (e.g., reading and recognizing a word).

Donald's analysis focuses on the cognitive and social sides of consciousness. He does not address the issue of how the brain can engender a phenomenal conscious state (the so-called hard problem; see Chap. 1). This lacuna notwithstanding, his picture remains a fascinating vision of the magnificent voyage that the human mind has made to develop its actual shape.

Essential Bibliography

- Donald M (1991) *Origin of the modern mind: three stages in the evolution of culture and cognition*. Harvard University Press, Cambridge, MA.
The book is a fascinating journey into the origin of the human mind. Donald tries to reconstruct the history of cognitive inventions that led humanity to its modern way of thinking. He thoroughly describes three phases in the evolution of human cognition, each of which characterized by a different system of representation that testified the passage from a mental framework to a more accomplished one.
- Donald M (2001) *A mind so rare: the evolution of human consciousness*. W W Norton & Company Incorporated, New York.
This book deals more specifically with the topic of consciousness than the *Origin of the Modern Mind*. Donald identifies three levels of brain conscious elaboration that are at the basis of animal behavior. Moreover, the three stages of human cognition are integrated in a broader view about how biological and cultural forces can interact with each other in order to originate consciousness.

The Three Worlds and Their Interaction

Human beings are irreplaceable; and in being irreplaceable they are clearly very different from machines. They are capable of enjoying life, and they are capable of suffering, and of facing death consciously.

They are selves; they are ends in themselves, as Kant said.

(The Self and Its Brain)

Sir John Carew Eccles was one of the most important neurophysiologists of the twentieth century. He was the 1963 Nobel Laureate in Physiology or Medicine, together with Andrew Huxley and Alan Lloyd Hodgkin, for his work on synapses. He was born on 27 January 1903 in Melbourne, Australia, and died on 2 May 1997 in Tenero-Contra, Switzerland. He studied medicine at the University of Melbourne, where he received his MD in 1925. He then moved to Magdalen College, Oxford University, from which he received his DPhil in 1929, under the supervision of Sir Charles Sherrington, who was, in turn, the 1932 Nobel Laureate in Physiology or Medicine. Eccles contributed in important developments within neuroscience by performing key experiments in synaptic transmission. In 1937, Eccles returned to Australia. After the war, he became a professor at the University of Otago, New Zealand, and from 1952 to 1962, he had a chair at the John Curtin School of Medical Research of the Australian National University. In 1966, he moved to the United States to work at the Institute for Biomedical Research in Chicago. He was then appointed as professor at the University of Buffalo from 1968 to 1975. Within the field of philosophy of mind, he is renowned for having held an original interactionist position together with philosopher Karl Popper.

Austrian-British philosopher Sir Karl Raimund Popper was one of the greatest thinkers of the twentieth century. He made astounding contributions to the philosophy of science as well as to the philosophy of politics. Within philosophy of science, Popper mainly reflected on the problem of demarcation between scientific and nonscientific theories, advocating the criterion of falsifiability in order to distinguish between them. In a nutshell, a certain hypothesis is scientific only if it makes predictions that can be falsified by experiments. This has been such a powerful idea that since its proposal it has never ceased to influence thinkers and stir debates. Within philosophy of politics, Popper was a champion of democracy, defined as “open society,” liberal ideas, and critical thinking.

Popper was born on 28 July 1902 in Vienna, Austria, and died on 17 September 1994 in London, England. Popper obtained a primary school teaching diploma in

1925 and a PhD in philosophy in 1928 from the University of Vienna. In 1937, he took up a position teaching philosophy at the University of Canterbury, New Zealand. He held this position until the end of World War II, and he subsequently taught logic and scientific method at the London School of Economics, University of London (1949–1969).

The interactionist position held by Popper and Eccles stands as peculiar in the panorama of philosophical and scientific theories about the nature of mind. Interactionism is a type of dualism claiming that mind and body are two different entities in interaction with each other. Thus, even though they belong to different realms, they are able to exert influences on one another. The most famous example of an interactionist theory of mind is the classical version of dualism developed by Descartes (see Chap. 6). Popper and Eccles' interactionism, however, is not a dualistic but a *pluralistic* theory of mind. In fact, it is not based on the metaphysical assumption that there be two kinds of realities (the mental and the physical) but on the hypothesis that nature is divided in at least three different ontological domains (or worlds) in reciprocal interaction.

The first domain or world 1 consists of physical bodies, such as stones, stars, water, plants, animals, radiation, and every other forms of physical energy. This world can be further divided into the realms of living things and nonliving physical objects, even though this distinction, as Popper points out, is not strictly defined.

The second domain or world 2 consists of mental and psychological phenomena. It is the world of feelings, emotions, thoughts, perceptions, judgments, decisions, observations, and every other cognitive process and subjective experience. World 2 is the most important for human beings and can be further divided in other several categories, such as conscious experiences, dreams, subconscious processes, human and animal minds, etc.

The third domain or world 3 consists of the products of human intelligence and ingenuity, such as language, stories, myths, geometric theorems, scientific theories, mathematical equations, symphonies, novels, artistic works, architectural projects, etc. This world can be subdivided in as many categories as there are disciplines of human knowledge.

Most entities can belong for some aspects to one world and for other aspects to another world. A book, for instance, is a physical object of world 1 but also an item of world 3 if considered under the aspect of its information. Similarly, Leonardo's Mona Lisa is an object belonging to world 1 as well as a painting belonging to world 3; Chopin's Polonaise Op. 53 no. 2 in A flat major for piano is a work belonging to world 3 as well as a pattern of auditory perceptions and feelings in the listeners, which belong to world 2 (Fig. 21.1). In turn, an emotion such as sadness is an entity of world 2 but also an entity of world 1 if considered under the aspect of its neural correlates. In a sense, many objects of world 1 *embody* or *physically realize* objects of worlds 2 and 3.

This embodiment or physical realization is literally the manifestation of the interaction between the worlds. In particular, the world of the contents of thought (world 3) can have causal effects on the world of physical objects (world 1). For

Fig. 21.1 Frédéric Chopin (1810–1849), autographed score of Polonaise Op. 53 no. 2 in A flat major for piano (1842), Heineman Music Collection, Pierpont Morgan Library Dept. of Music Manuscripts and Books, New York, US



instance, scientific conjectures and theories are great instruments in order to change world 1. The interaction between world 3 and world 1 occurs within the brain at the level of the world of mental states (world 2). In particular, Eccles postulated the existence of mental units (called *psychons*), which are associated with analogous cerebral units (called *dendrons*) formed by set of dendrites (neuronal extensions receiving impulses from other neurons and transmitting them to the cell body). Eccles' proposal is that the interaction between mind and body occurs at the level of microscopic components formed by the associations of psychons with dendrons and can be accounted for in terms of quantum mechanics. As we shall see, the idea that the rules of quantum mechanics might play a pivotal role in the production of consciousness has also been advanced by other thinkers (see Chap. 27).

According to Eccles, psychons are the very unitary experiences of consciousness and could originate in living organisms only when the cerebral cortex developed a sufficiently complex structure of dendrites capable to organize dendrons. This higher-order brain architecture appeared only with the advent of mammals, whose cerebral cortex eventually evolved with a propensity for relating to a world other than the physical one. Thus, the evolutionary origin of consciousness can be dated at least as early as 200 million years ago, when gleams of consciousness could first appear in the mammalian insectivores. On the contrary, the reptilian brain would not have reached a sufficient degree of complexity to generate consciousness, as well as fish, even though it had the potential to evolve into a primitive mammalian brain and into the brain of birds (a region of which called *pallium* could have analogous functions to those of the mammalian cortex).

Although Popper and Eccles' interactionist approach of the three worlds has the merits to place consciousness in an evolutionary perspective and to stress the importance of neuroanatomy in order to understand the processing of the conscious mind, a few unaddressed issues stand out. Theoretical efforts have been mainly directed at the clarification of the connection modality between world 1 and world

2, while no explanation has been proposed as to how abstract entities or contents of thoughts could be linked to mental phenomena. In other worlds, the connection between world 3 and world 2 remains somewhat unclear. Moreover, if we were inclined to accept that world 2 (mental entities) – and through it world 3 (abstract objects) – can exert causal effects upon physical things, then we could face a case of *causal overdetermination*, that is, a case in which the same effect is simultaneously brought about by both a physical cause and a mental/abstract cause (see Chap. 8 for a discussion about this problem). Thus, the acceptance of the possibility of three autonomous worlds in reciprocal causal interaction seems to lead to an unjustified proliferation of causes.

Essential Bibliography

- Popper KR, Eccles JC (1977) *The self and its brain: an argument for interactionism*. Springer, Berlin; reprinted in 2012, Springer London, Limited.
The authors argue that the mind-body problem can be adequately accounted for within a pluralistic framework, according to which the world of material objects (world 1), the world of mental states (world 2), and the world of abstract entities (world 3) reciprocally interact. The book is divided in three sections. In the first section, Popper discusses the philosophical issues surrounding the mind-body problem. In the second section, Eccles analyzes the mind from a neuroscientific standpoint. The third section consists of twelve dialogues in which the two authors come to terms with some conflicting opinions.
- Eccles JC (1989) *Evolution of the brain: creation of the self*. Routledge, London. Within an evolutionary (but non-Darwinian) perspective, the author tells the story of how the human mind came to be. The development of the human brain is intriguingly narrated from the point of view of a detective mystery.
- Popper KR (1994) *Knowledge and the mind-body problem: in defence of interaction*. Routledge, London.
The book is based on lectures given by Popper at Emory University in 1969. It deals with the nature of objective knowledge, which is related to the existence of a world of abstract entities (world 3) and addresses the mind-body problem in the light of an interaction between a world of physical states (world 1) and a world of mental states (world 2).

The Reentrant Dynamic Core

My thesis is that the evolution of a reentrant thalamocortical system capable of giving rise to the dynamic core allowed the integration of vastly increased complexes of sensorimotor inputs. Animals having such a core were therefore capable of refined discriminations. Qualia are just those discriminations, each entailed by a different core state.
(Second Nature: Brain Science and Human Knowledge)

Gerald Edelman (born on 1 July 1929 in New York, died on 17 May 2014 in La Jolla, San Diego) received an MD in 1954 from the University of Pennsylvania School of Medicine. Following an impressive academic career as an original researcher, in 1972 he won the Nobel Prize in Physiology or Medicine for his work on the immune system. After the Nobel Prize award, he pursued research in the regulation of cellular processes, focusing on cell-to-cell interactions in the embryonic development and in the formation and function of the brain and neural networks. In 1992, he moved to California and became Professor of Neurobiology at the *Scripps Research Institute*. He also founded the *Neurosciences Institute* in San Diego, which flourished under his direction. He has written popular books on his theory about the nature of consciousness, one of which coauthored with Giulio Tononi (see Chap. 28), with whom he contributed to develop an intriguing hypothesis on how the brain can consciously elaborate information.

Edelman's theoretical framework for the study of consciousness is entirely biologic and naturalistic. Specifically, consciousness is considered as a natural property of the living human brain and should, accordingly, find its cause in neurophysiology. For this reason, Edelman has repeatedly pointed out that the brain is not to be equated to a computer. In fact, in order to confront with an ever-changing reality, the brain must rely on a flexible and plastic organization capable of sorting out ambiguous information in a coherent fashion. This result is achieved by the cerebral adaptive structure, which is formed through processes of selection upon variation. By contrast, computing machines – which are not based on these processes but rather on more fixed and less ambiguous instructions – struggle to capture the intrinsic variability of natural phenomena.

According to Edelman, two main interactive processes regulate the development of the brain. The first process, occurring especially in the embryo and during

the postnatal period, is the formation of neuronal groups of variable size and structure, in which numerous adjacent neurons tend to be strongly interconnected. The second process, which is long life and accounts for the several differences across individual brains, is the strengthening and progressive modification of synaptic junctions between neurons. This plastic remodeling, which takes place during daily activities, literally shapes the neuronal groups by selecting those which yield the most adaptive behavior. In other words, only the neuronal groups that are able to respond more efficiently to the environmental demands are destined to survive within the brain assembly. In light of this selective feature, Edelman calls his approach *neural Darwinism* and refers to his dynamic model as the *theory of neuronal group selection*.

Edelman's theory of neuronal group selection accounts for the development of adaptive behavior by virtue of three mechanisms: (1) developmental variation and selection, (2) experiential selection, and (3) reentrant signaling. Each mechanism acts within and between neuronal groups by introducing adaptive changes to their synaptic strength, so that the responses of the group as a whole can be enhanced or decreased. Thus, even though neuronal groups mainly arise from contiguous anatomical connections, groups themselves are dynamic clusters of neuronal patterns which are affected by both synaptic changes and informational inputs to the individual groups.

The first mechanism of selection is characterized by genetic and epigenetic regulation mechanisms for cell division, adhesion, migration, death, and pruning, including neurite extension and retraction. All these processes influence cell motion and association, leading to complex patterns of axons and dendrites' ramifications. Highly differentiated anatomical brain regions arise, as well as large numbers of variant neuronal groups and circuitries. As a result, a structurally diverse primary repertoire of local neuronal patterns is created by differential reproduction.

The second mechanism is characterized by postnatal synaptic modifications driven by experience. Selection occurs within populations of synapses, which are strengthened or weakened without major changes in the anatomical architecture of neuronal groups. This process, in turn, creates a diverse secondary repertoire of behaviors by differential amplification. In fact, the functioning of neuronal groups continues to be dynamically selected in response to particular signal patterns. Signals of a similar type increase activity in previously selected circuits and favor specific neuronal group functions over others.

The third mechanism is characterized by a higher-order selection process which is called reentry. Reentry can be described as ongoing parallel and bidirectional signals traveling between separate neuronal groups in a recursive fashion. As a result, reentrant connections form a dynamic and largely distributed functional process, which can account for the coordination of different neuronal patterns. The main characteristic of reentry is that it occurs in parallel; it is therefore different from feedback, which involves a serial transmission along a single pathway. Reentrant connections can also explain the so-called binding problem, that is, how functionally segregated cortical regions, which are specialized for each

sensory modality, can be bound together in order to provide a perceptually coherent picture of reality.

Edelman's theory of neuronal groups selection leads him to suggest that conscious experience does not derive from the activity of a single brain location or neuronal type but rather is the product of dynamic interactions between widely distributed groups of neurons. Edelman points out that these mutual connections are pivotal to understanding the properties of consciousness, which can be classified in three categories: general, informational, and subjective. The general features of consciousness are as follows: (1) conscious states are unitary and integrated; (2) they can be extremely differentiated; (3) they are experienced along a temporal order; (4) they are able to bind different sensory modalities; and (5) they have constructive properties (such as phenomena of filling in). The informational features of consciousness are as follows: (1) conscious states have intentionality and wide-ranging contents; (2) they show both accessibility and associability; (3) they are characterized by both center and fringe features; and (4) they are modulated by the attentional process. The subjective features of consciousness are as follows: (1) conscious states reflect phenomenal experiences, qualia, and feelings; (2) they are situated in the world; and (3) they are related to feelings of familiarity or unfamiliarity.

Therefore, the neural systems that underlie consciousness can integrate many informational inputs in a short time, so as to enable higher-order discriminations within a multidimensional space. According to Edelman, these sophisticated discriminations are the phenomenal experiences or qualia that constitute the subjective nature of consciousness (similar ideas can be found in the theory of consciousness put forward by Tononi, see Chap. 28). Thus, consciousness emerges in a system which is able to maintain both sufficiently large repertoires of different circuits and a massive reentrant set of connections. This complex organization allows the interaction between many heterogeneous components so as to form larger assemblies which, in turn, yield new integrated functions. According to Edelman, the thalamocortical system is an essential functional cluster for the generation of consciousness. Its activity is in fact mainly responsible for the production of a unitary experiential scene, in which the different contents of consciousness coming from functionally segregated brain areas are dynamically coalesced.

Edelman calls the functional activity of the thalamocortical system the *reentrant dynamic core*. Higher-order discriminations or phenomenal experiences (*qualia*) rely on the functionality of this dynamic core. The dynamic nature of the core implies that neuronal groups which are involved in the core activity at a certain moment can change. In other words, neuronal groups that previously were not in the core can enter it successively, whereas others that were previously incorporated in the reentrant process can leave it. The dynamic core gathers the signals coming from the world and the body and converts them in a *phenomenal transform*. This process of conversion is what we experience as phenomenal consciousness (*qualia*).

According to Edelman, the *phenomenal transform* has no causal power at all. *Qualia* or conscious experiences are entailed by material processes but are not themselves material entities; they are, instead, by-products which have no part in the

Fig. 22.1 Flemish School, *Plato's Cave* (sixteenth century), Musée de la Chartreuse, Douai, France (Image cropped by authors)



causal chain of events of the physical world. Within this picture, consciousness appears to be no more than an epiphenomenal consequence of the neural activity of the dynamic core. By contrast, we are wrongly used to think that our mental states are real and causally effective, like the chained men in Plato's cave, who misleadingly believe that the shadows they see on the wall of the cave belong to reality (Fig. 22.1).

Edelman admits that phenomenal conscious states are not meaningless and unnecessary, because they are highly informative (i.e., unitary and, at the same time, differentiated): they are therefore quite useful to concisely convey a huge amount of information. However, Edelman reminds us that our habit of speaking about conscious experiences can be nothing more than a matter of linguistic convenience.

Edelman's epiphenomenalism seems to be one of the most arguable sides of his theory. In fact, Edelman claims that the relationship between the neural activity of the dynamic core and consciousness is strictly necessary. Specific activities of the nervous system necessarily give rise to particular conscious experiences, which in turn cannot emerge without a specific underlying brain activity pattern. If this is the case, then there is no reason to assume that conscious mental events and physical processes within the dynamic core are to be considered as distinct. For theoretical purposes, it would be much easier to maintain that conscious mental properties are *identical* to certain physical activities occurring in the dynamic core. This picture would also be more consistent with the claim that conscious phenomena are highly informative, meaningful, and intentional, as well as with the neuropsychological evidence showing that impairments in the conscious representation of the body heavily affect the individual's behavior.

Moreover, Edelman points out that the dynamic core is strongly influenced by diffuse ascending limbic value systems, which continuously contribute to regulate synaptic activity regarding memory, attention, and perceptual categorization on the basis of species-specific responses and past experiences. In turn, homeostatic systems in other deep brain regions help to distinguish between self and non-self

categorization. All these interactions allow the emergence of *primary consciousness* (also called by Edelman the *remembered present*), that is, a multimodal scene of the present constructed out of perceptual and motor signals. In a subsequent stage of evolution, animals with primitive semantic capabilities can develop a higher-order consciousness or *secondary consciousness*. Nonetheless, episodic memory and language eventually allow the generation of a full-blown sense of subjectivity only in human beings. Epiphenomenal consciousness does not seem to fit well in this picture, given that secondary consciousness gives significant advantage over other animals with the ability to develop just primary consciousness. Primary consciousness, in turn, gives advantage over unconscious living creatures. Consciousness, therefore, appears to play an active and causal role in the battle for survival, by prompting specific and successful behaviors.

Edelman's theory of neuronal group selection is one of the most interesting hypotheses of how the brain develops its functions. It can also offer a promising biological framework for research, as well as reinforcing the idea that consciousness is a natural phenomenon among several others, albeit fascinating and extremely difficult to tackle by scientific exploration.

Essential Bibliography

- Edelman GM (1989) *The remembered present: a biological theory of consciousness*. Basic Books, New York.
This book offers a thorough description of Edelman's theory of neuronal group selection and proposes a biological and naturalistic framework for the understanding of consciousness. The distinction between primary and secondary consciousness is clearly explained, as well as the central notion of reentry in brain structure and dynamics.
- Edelman GM (1992) *Bright air, brilliant fire: on the matter of the mind*. Basic Books, New York.
In this book, Edelman criticizes the computational theory of mind as well as the claim that mental phenomena do not require an explanation in biological terms. By contrast, Edelman holds that both the brain and its functional product, the conscious mind, can only be understood within an evolutionary perspective.
- Edelman GM (2007) *Second nature: brain science and human knowledge*. Yale University Press, New Haven, Connecticut.
This book further develops Edelman's position that consciousness is destined to lose its mystery by virtue of its scientific study and to be eventually included in the natural order. At the same time, however, Edelman addresses the limits of the scientific approach with regard to creativity and acquisition of knowledge, which are inextricably linked to each individual's irreversible and unique evolutionary history.

Consciousness As Recursive Intentionality

*I have written of consciousness as a surface feature of the brain
and so I think it is, but you will see now that I am suggesting
it is a very special sort of surface feature.
For what consciousness actually is, is a feature
not of the whole brain but of this added self-reflective loop.
(The Mind Made Flesh)*

Nicholas Keynes Humphrey (born on 27 March 1943) is a British psychologist with wide-ranging research interests. Some of his most relevant contributions focus on the evolution of human intelligence and consciousness. Humphrey was educated at Trinity College, Cambridge, from which he gained both his BA in Psychology and Physiology and his PhD in Psychology. His doctoral research, supervised by Lawrence Weiskrantz, focused on the neuropsychology of vision in primates. His early research concerned the study of blindsight after brain damage in monkeys. He first demonstrated this phenomenon in a monkey called Helen. In 1971, he also spent a 3-month period at Dian Fossey's Gorilla Research Centre in Rwanda. Humphrey has been Lecturer in Psychology at Oxford; Assistant Director of the Subdepartment of Animal Behaviour at Cambridge; Senior Research Fellow in Parapsychology at Cambridge; Professor of Psychology at the New School for Social Research, New York; and School Professor at the London School of Economics. He has also worked on a number of TV and radio documentaries.

Humphrey's theory of consciousness develops within neuroscience but at the same time leads us to redefine some common ideas derived from the use of our ordinary language. Humphrey starts to claim that consciousness and brain activity are not to be considered as *incommensurable* phenomena – that is, phenomena that do not have common standard of measurement – but rather as the sides of the following identity equation: *mental state, m = brain state, b*. The sort of identity claimed by Humphrey is not a type of materialistic identity between mental and brain states but a functional one. Therefore, even if the two terms of the equation seem to belong to different dimensions, Humphrey suggests that cognitive science will be able to describe both mental states and brain processes in computational or functional terms, that is, in terms of rules connecting inputs to outputs. This functional description will put the two terms of the equation on the same ground, thus

wiping out the illusion of their incommensurability. In the meantime, Humphrey aims to give theoretical support to this research strategy.

While the functional analysis is relatively “easy” for the so-called *intentional* mental states (i.e., mental states directed to objects) – such as memories, perceptions, desires, thoughts, as well as the underlying processes of remembering, perceiving, wanting, and thinking – subjective sensations that involve phenomenal consciousness seem instead to be unapproachable using functional tools. The experience of seeing the blue sky, the taste of wine, and the pain of a needle piercing a finger, all these sensations seem to have an additional ingredient which cannot be expressed in functional terms. As we have seen, these fleeting subjective experiences are called qualia by philosophers, although Humphrey chooses to refer to them as *sensory phantasms*, following seventeenth-century scientist Isaac Newton’s terminology.

With respect to qualia, the terms of the identity equation are *phantasm*, $p = \text{brain state}$, b . Given that qualia cannot be functionally analyzable, according to Humphrey, there are three different ways to treat this equation. The first way is to consider the equation as false. In this case, there would not be a relation of identity between mental and brain processes but only a sort of correlation. This correlation, nonetheless, needs to be explained. The second way is to consider the equation as true but accept at the same time that a scientific explanation of the identity between mind and brain is beyond human reach. We will never know why the identity holds true; this will always be an unsolvable mystery. The third way, which is the option chosen by Humphrey, is to consider the equation as true and nonetheless persist in the effort to find a solution for it. In this case, the only way to proceed is to reconceive of the terms of the identity equation in order to make them line up.

According to Humphrey, the best strategy is to operate on both sides of the equation, so as to redefine our concepts of sensory phantasms and brain states until they eventually match up. Starting with the left side, Humphrey endorses a subtle conceptual distinction between *perception* and *sensation*, which was first claimed by eighteenth-century Scottish philosopher Thomas Reid. According to Reid, human beings use their senses in two different ways: they use them *to perceive* the external world, on the one hand, and *to feel* the inner world, on the other. Through perception, humans acquire the concept or belief of the existence of external objects. Through sensation, they have the feeling of how their body is affected by the external objects. In the former activity, sensory stimulation provides an objectively and affectively neutral representation of what is out there; in the latter, sensory stimulation provides a subjective and affect-laden representation of what is currently going on inside the perceiver’s body. In other words, perception has to do with judgments about facts happening in the external world, whereas sensation has to do with feelings about what is happening to the sentient being.

Thus, even though both sensory activities involve bodily stimulation, their outcomes and nature are different and need to be maintained as distinct. However,

laymen as well as many philosophers do not generally distinguish the two processes and assume that perceptual judgments have a sort of intrinsic phenomenal content. Some philosophers of mind, who take the issue about the existence of qualia seriously, tend to think that there is something it is like to perceive, say, the red sunset or the shape of a car, the smell of garlic, the taste of chocolate, etc. Humphrey dubs the supposed phenomenal aspect of perception *pseudo-sensory phenomenology*. In truth, perceptions have no phenomenal content; phenomenal content is all on the side of sensation. However, according to Humphrey, sensations have no phenomenal content in the same way as perceptions have objects. Following Thomas Reid again, he claims that language misleads us when we talk of feeling or having certain sensations as if these sensations were objects or sense data coming from our sensory processing. On the contrary, sensations are no more to be considered as the objects of sensing or feeling than thoughts the objects of thinking, intentions the objects of intending, and volitions the objects of willing. This is because sensing or feeling is not a passive state but rather a form of active engagement with what is currently affecting the body.

Humphrey identifies five defining properties of a conscious sensory experience: (1) ownership, (2) bodily location, (3) presentness, (4) qualitative modality, and (5) phenomenal immediacy. First, a conscious sensory experience always belongs to the subject who is having it; in other words, every sentient being is the one and only author of its sensations. Second, any conscious sensory experience is indexical, that is, dependent on a certain context, and involves a specific part of the perceiver's body. Third, any conscious sensory experience always occurs at the present tense; it is in existence just here and now for the time being. Fourth, any conscious sensory experience always presents a certain qualitative flavor; it involves a distinct sensory modality (visual, olfactory, tactile, etc.). Fifth, any conscious sensory experience is phenomenally immediate, that is, it is intrinsically and directly phenomenal in itself. This last property implies that the other four properties are self-disclosing: when a conscious sensory experience occurs, the perceiver is directly aware that he or she is the only author of that sensation, as well as that the sensation is localized or contextualized, is happening in the present moment, and has a peculiar qualitative feature.

According to Humphrey, when we feel pain or we see a red sunset, what it is really happening is that we *are paining* or we *are feeling redly* about a certain part of our visual field. As a result, phenomenal consciousness is a direct and unmediated feature of a modality-specific way of feeling. In other words, for a person to be the conscious subject of a sensation simply means for him or her to be in the appropriate sensory activity, so that for him or her to have a red sensation is to do the "redding," to have pain is to do the "paining," etc. Humphrey proposes to call this peculiar activity *sentition*, similarly to the other mental processes of *volition* and *cognition*. He therefore explains the experience of what it is like to be something as being the particular mental state which is able to take itself as its own intentional object, in a continuous process of *self-resonance* or *recursive intentionality*.

Fig. 23.1 Johannes Gump (1627 – died after 1646), *Self-portrait*, oil on panel (1646), Uffizi, Florence, Italy (Image cropped by authors)



Since the mind involves a process of actively doing something about something, then also the brain must involve a process of actively doing something about something. In other words, if the identity equation holds, the brain will do exactly the same thing the mind does. Thus, to match the right side of the equation to the left one, we need to find within the brain a recursive or self-resonant process, which is able to be the subject of itself so as to generate a sort of self-sustaining loop or self-reverberating circuit, as it could be artistically represented in the ingenious self-portrait made by seventeenth-century Austrian artist Johannes Gump (Fig. 23.1).

The concept of “loop” has been analyzed and considered to have a central role in creating higher-order cognitive processes within both the philosophical and scientific frameworks. Likewise, from a neuroscientific perspective, it has been shown that reentrant circuitries are thought to be essential features of conscious brain activity, and the thalamocortical system appears to be the best candidate for elaborating information in a self-reverberating manner (see Chaps. 22 and 28).

Humphrey has also discussed the role of phenomenal consciousness and suggests that its evolutionary purpose “may not be to *enable* us to do something we *could not* do otherwise, but rather to *encourage* us to do something we *would not* do otherwise: to make us *take an interest* in things that otherwise would not interest us, or to *mind* about things we otherwise would not mind about, or to *set ourselves goals* we otherwise would not set” (2008). In other words, consciousness would exist to make life more precious or worth living. This seems to be a very simple and elegant solution for explaining the role of consciousness. It could be argued that consciousness is equally responsible for what is good and for what is bad in life. Consciousness can make us fill with wonder for nature but also make us fall in the abyss of despair. Moreover, from an ecological perspective, consciousness might serve a plurality of roles, which encompass phenomenological, cognitive, and social domains, thus offering wide scope for the refinement of Humphrey’s explanatory model.

Essential Bibliography

- Humphrey N (1992) *A history of the mind: evolution and the birth of consciousness*. Chatto & Windus, London.
The author narrates a fascinating tale about how and why consciousness has evolved. The distinction between the concepts of “perception” and “sensation” is thoroughly analyzed as well as the mechanism of the self-sustaining loop.
- Humphrey N (2006) *Seeing red: a study in consciousness*. Harvard University Press, Cambridge, MA.
With a clear and concise style, this book summarizes the author’s theory of the origin and nature of consciousness, tracing back its evolution to our primordial ancestors’ expressions of liking and disgust. It further develops the idea that sensations are not things that happen to us but rather things that we actively do.
- Humphrey N (2011) *Soul dust: the magic of consciousness*. Princeton University Press, Princeton.
The author discusses in detail the evolutionary purpose of consciousness and suggests that it plays a central role in human life. The conclusion is simple and straightforward: consciousness is just the most special ingredient of our lives.

The Bicameral Mind

Suppose you ask a flashlight in a completely dark room to turn itself on and to look around and see if there was any light – the flashlight as it looked around would of course see light everywhere and come to the conclusion that the room was brilliantly lit when in fact it was mostly just the opposite. So with consciousness. We have an illusion that it is all mentality.
(Consciousness and the Voices of the Mind)

American psychologist Julian Jaynes authored one of the most thought-provoking and debated theories about the origin of the conscious mind. He was born on 27 February 1920 in West Newton, Massachusetts, and died on 21 November 1997. He studied as an undergraduate at Harvard and McGill Universities and received his master and doctoral degrees from Yale. He then became a lecturer in Psychology at Princeton University from 1966 to 1990. At the beginning of his career, Jaynes pursued research in the field of animal behavior, as his theoretical approach was to investigate the evolution of consciousness by studying learning and brain function in various species, from the protozoa to worms, reptiles, and cats. Finding this approach inadequate, he undertook a painstaking analysis through historical texts and archeological data in order to investigate the use of language and metaphor related to the theme of consciousness. This in-depth research culminated in 1976, when Jaynes published his landmark book – *The Origin of Consciousness in the Breakdown of the Bicameral Mind* – which proposed a revolutionary theoretical model for the generation and historical development of consciousness.

Jaynes' theoretical analysis starts by asking what consciousness is and what it is not. First, Jaynes points out that the nervous system is able to implement a large amount of autonomic processes without the help of consciousness. Indeed, it seems that many brain functions totally escape our conscious control. For instance, all the perceptual constancies with regard to shape, color, size, and brightness of objects are accomplished by the brain without any intervention from consciousness. Similarly, we are not always aware of how we sit, walk, and move. Even in the production of speech, consciousness does not seem to be determinant. In fact, we are not conscious of the selection of words until they are said. In a sense, words pop into the conscious mind like a sudden flash in the night sky. What we are aware of, instead, are intentions of certain meanings, which Jaynes calls *struc-tions*. Thus, it is as if the brain chooses the words after receiving the structions of

intended meanings. A similar remark was made by American writer Edward Morgan Forster when he asked in *Aspects of the Novel* (1927) “How do I know what I think until I see what I say?”

Another important point highlighted by Jaynes is that the contents of conscious experience do not provide an exact copy of reality. By contrast, the conscious mind continuously constructs and reconstructs its own world. Furthermore, consciousness does not seem to be essential for simple learning tasks, such as the acquisition of conditioned responses and new motor routines. Specifically, Jaynes claims that the conscious mind plays a role in decisions as to what to learn and in verbalizing features of a learning task, but it is not active in the learning process itself, which runs through automatic pathways. Consequently, consciousness seems to be even inessential for thinking and reasoning. A similar statement was already pointed out by French physicist Henri Poincaré, who described how suddenly ideas can creatively pop into his head after prolonged unconscious ruminations.

Jaynes concludes, therefore, that it is an illusion to think that consciousness influences everything we do. And given that a number of brain activities and behaviors can be carried out without consciousness, Jaynes argues that consciousness could have had a later development with respect to the other brain functions. In other words, he suggests to seriously consider the possibility that there could have been in the past human beings with unconscious minds but who were able to do most of the things we do, such as perceiving, speaking, and solving problems. Still, something in the functionality of the brain subsequently changed, leading to the development of the modern mind. According to Jaynes, what emerged was not a novel biological property but a radical different way of using language.

Jaynes notices that whenever we use words to refer to mental events, in fact we speak metaphorically of something that is part of a behavioral world. According to his model, adjectives and expressions, which are used to describe physical events in real space, are analogically applied to describe mental events in a virtual mind-space. In light of this, mind can be depicted as “quick,” “slow,” and “sharp,” just like somebody can be described as being “broad minded,” “strong minded,” “narrow minded,” etc. Similarly, by using spatial categories, we can say that something is “beyond” or at the “front”/“back” of the mind. According to Jaynes, the use of metaphors literally creates a landscape within the mind, in which consciousness analogically emerges.

Jaynes identifies two components in the process of generating a metaphor: the *metaphrand* (the object expressed in words) and the *metaphier* (the term prompted by the struction). For example, we can exclaim “I see” when we suddenly get the solution of a problem. In this case, the expression “I see” is the metaphier, and the mental event that has grasped the solution is the metaphrand. Metaphiers are usually linked to what Jaynes calls *paraphiers* (meaningful associations with other words, concepts, and experiences, which are attributes shared with the metaphrand through the metaphier). In turn, paraphiers project back to the metaphrand so as to generate a further element called *paraphrand* (a new linguistic entity which is united to the metaphrand in order to create within the mental

space an analog of the object described by the metaphor). This linguistic creative process generates the following loop:



To clarify, let us analyze the metaphor “clouds cry tears of rain when they miss the sea.” In this expression, the metaphrands are the things described in relation with each other (clouds, rain, and sea); the metaphiers are the terms that convey the metaphor (the act of crying and tears); the paraphiers are the qualitative associations and implications of crying (the sadness for missing the sea); the paraphrand, finally, is the mental product of the new image derived from the relationships between the clouds, tears, and rain within the virtual mental space.

According to Jaynes, the emergence of this mental relational space is the primary feature of consciousness. This space, in turn, can be analogically introspected by a further metaphorical construction, which Jaynes calls the *analog “I.”* The analog “I” is the second most relevant feature of consciousness and corresponds to the projection of the body into the mental space: just as the bodily “I” can move in the physical space, so the analog “I” can move inside the mind-space. This introspection leads to the third fundamental feature of consciousness, which is *narratization*. In fact, consciousness constantly makes stories by weaving together different things and events in a logical temporal sequence. Similarly to the body, which travels through the physical world in spatial successions, the analog “I” travels through the mental space in a *spatialized time*, thereby assigning a before and an after to any event.

Thus, according to Jaynes’ theory, consciousness appears to be an analogical construction of the world, intimately bound up with volition and decision but also essentially based on linguistic reentrant processes. Following this line of thought, the manifestation of language must have preceded consciousness and made possible a structure for its evolutionary emergence. Therefore, according to Jaynes, we should not ask for explanations of consciousness in biological or neurophysiological terms, but we should rather ask when in human history the use of language developed so as to create an analog “I,” which was able to self-visualized and be the subject of a narrative story within a mental space (for a further development of the idea that consciousness is intimately linked to the narrative power of mind, see Chap. 5).

On the grounds of an audacious philological interpretation of ancient literary texts, Jaynes suggests that the *Iliad* could provide evidence for a time in which human beings had already developed language but were still unconscious. Intriguingly, in the ancient traditions collected in the Iliadic text, there are no single words which can translate our concepts of “consciousness,” “mind,” “soul,” and even “body.” There are instead several terms – called *preconscious hypostases* by Jaynes – which refer to physiological processes related to mental life. Examples are

Fig. 24.1 Peter Paul Rubens (1577–1640) *Achilles slays Hector* (seventeenth century), oil on canvas, Museum Boijmans Van Beuningen, Rotterdam, Netherlands (Image cropped by authors)



words like “psyche” (the living breath, which departs from the body at the moment of death), “thumos” (either the blowing breath or the flowing blood), and “phren” (frequently in the plural form “phrenes,” which could have referred to the inflating lungs). These lexical oddities led Jaynes to think that the heroes of the *Iliad* were like noble “automata,” unable to introspecting or reminiscing in order to make decisions. For these people important decisions were taken in form of verbal hallucinations, which were considered as having a divine origin. This is why Homeric heroes are often described in the *Iliad* as accompanied by deities who suggest them what to do and guide their behavior (see Fig. 24.1).

Jaynes suggests that the absence of insight shown by the characters of the *Iliad* reflects a time when the human brain was organized according to a bicameral architecture. The right synthetic and metaphoric hemisphere was able to transmit hallucinatory verbal instruction to the left analytical and rational hemisphere (the interpreter), particularly in case of stressful situations. Human mentality was therefore divided in two parts, neither of which was conscious. A subsequent reorganization in the cognitive architecture of the brain would have led to a new way of using language and, as a result, to consciousness. Jaynes claims that this reorganization could have occurred around 1400–600 B.C., when the chaos of massive migrations generated by natural catastrophes, overpopulation, and the emergence and diffusion of writing could have caused the breakdown of the bicameral mind. Jaynes proposes this period both because the *Iliad* is a collection of more ancient oral poems assembled around 700 B.C. and because the lack of introspection in literary characters had already disappeared by the time the *Odyssey* was composed.

Jaynes claims that remnants of the ancient bicameral brain organization can still be found in the verbal or auditory hallucinations associated with hypnosis and schizophrenia. This hypothesis is consistent with the findings from recent neuroimaging studies, which identified the right temporal lobe as the source of auditory hallucinations in patients with schizophrenia. However, the great variety of these phenomena observed in both normal people and neuropsychiatric patients suggests that the stressful line of events proposed by Jaynes is too simplistic to provide a valid etiological ground. On the contrary, neurophysiological studies have given thus far little support for the bicameral architecture of the preconscious mind. At best, Jaynes' theory could help explain some cultural (software) rather than structural (hardware) developments of conscious processes.

Jaynes' idea that consciousness might emerge from the generative power of metaphors is arguable and, at the same time, opens the door to endless philosophical debates about the priority of language over consciousness. At present, the claim that consciousness ultimately depends on a metaphor-driven use of language is considered as fairly questionable by the majority of neuroscientists and philosophers of mind. Rather than being a fundamental prerequisite for consciousness, language appears to contribute to some higher faculties of the conscious mind, such as self-cognition and identity. On the other hand, Jaynes' multidisciplinary approach to the problem of consciousness has several merits: not only it stirred intriguing issues of discussion, but it also stressed the importance of establishing cross-disciplines bridges for the study of consciousness.

Essential Bibliography

- Jaynes J (1976) *The Origin of consciousness in the breakdown of the bicameral mind*. Houghton Mifflin, Boston.
In this seminal book, Jaynes describes his thought-provoking and innovative theory of consciousness, based on a multidisciplinary analysis of evidence from neuroscience, psychology, philosophy, philology, and archeology. Jaynes is able to weave a fascinating hypothesis of how consciousness could emerge around 300 years ago by virtue of a new linguistic development in the functionality of the human brain.
- Jaynes J (1992) *The Julian Jaynes collection*. Basic Books, New York.
This is an anthology of Jaynes' articles, essays, lectures, and interviews on the theme of the origin and development of consciousness. His theory is clarified and expanded so as to comprehend new aspects of human nature, such as the notion of self, dreams, emotion, and artistic creativity.

The Unified Conscious Mental Field

*The brain was evidently beginning the volitional process
in this voluntary act well before the activation
of the muscles that produced the movement.
My question then became: when does the conscious
wish or intention (to perform the act) appear?
(Do We Have Free Will?)*

Benjamin Libet (Chicago, Illinois, 12 April 1916–Davis, California, 23 July 2007) was a scientist who conducted pioneering studies on consciousness and free will. He studied physiology, and between 1945 and 1948, he worked as an assistant professor at the University of Chicago. During his academic career, he was a lecturer at the Albany Medical College, New York; research fellow in neurochemistry at the Institute of the Pennsylvania Hospital, Philadelphia; and instructor at the University of Pennsylvania Medical School. In 1956, he collaborated with Sir John Eccles at Canberra (Australia) and subsequently became member of the Center for Neuroscience at the University of California, San Francisco.

Benjamin Libet's interpretation of his experiments has become one of the most debated and controversial issue in the fields of consciousness studies and philosophy of mind. In the 1970s, he was involved in research on sensory thresholds, directed to investigate the degree of activation needed to generate artificial somatic sensations in specific brain regions. This issue promptly turned his interest to the topic of consciousness and the voluntariness of action. In particular, he pursued an empirical method to test the hypothesis, proposed by Eccles, that the conscious intention or the act of will ought to precede the *Bereitschaftspotential* (readiness potential) discovered by Hans Helmut Kornhuber and Lüder Deecke in 1965. In their experiments, Kornhuber and Deecke found that when subjects were asked to voluntarily move their wrist or fingers, the movement was preceded by a slow electrical change recordable on the scalp at the vertex. This readiness potential appeared around 1 s before the movement and was thought to be an indication that the plan to perform the action had been prepared. Libet's experiment inserted in this line of research with the aim to demonstrate whether or not the readiness potential can precede the conscious decision to act.

The commonsense view is that the conscious will should manifest *before* (or at least at the onset of) the readiness potential. In contrast, Libet was able to show with a simple yet elegant experimental setting that the conscious will appeared to *follow* the readiness potential onset. The conclusion he drew was devastating for the



Fig. 25.1 Antonio de Pereda y Salgado (1611–1678), *Still-life with a Pendulum* (1652), oil on canvas, Pushkin Museum of Fine Arts, Moscow, Russia

ordinary conception of free will and stirred a debate that is still extremely vivid. The ingenious experiment was designed as follows. Subjects were put in front of a clock similar to that painted in a still-life by Pereda y Salgado (Fig. 25.1) in which a spot of light revolved around the face of the screen at a speed approximately 25 times faster than the speed of the sweep-second hand of a common clock. Each second indicated on the clock screen, therefore, was equivalent to about 40 ms. Subjects could flex the wrist at any freely chosen time, but they were asked to maintain their gaze directed to the clock center and note the position of the light spot when they became aware of the intention to move. This reported time (W) was associated with the first awareness of the wish to act. In turn, each readiness potential was obtained as an average of electroencephalographic recordings in 40 trials (Fig. 25.1).

In the groups of subjects in which the movement was freely expressed, the readiness potential preceded muscular activity by about 550 ms, whereas in the groups of subjects who reported some preplanning of the movement, the readiness potential preceded muscular activity by about 1,050 ms. In both groups, however, the reported timing of W (the conscious wish to act) was about 200–150 ms before the muscular activation. From these results, Libet drew the conclusion that the brain process which is involved in preparing the voluntary action began unconsciously at

least about 400–350 ms before the conscious will to move. It seemed, therefore, that the brain was to organize the movement well before the person could consciously recognize the wish to act. Subsequently, Libet asked whether there could still be a role played by the conscious will in the performance of a voluntary act. And his answer was positive. In fact, although the readiness potential precedes W, the interval of 200–150 ms between W and the muscular activation would allow enough time to consciously decide whether to accomplish the movement or to abort it. Therefore, in a sense, consciousness does influence the outcome of the volitional process by either approving or stopping it, even though the onset of the voluntary action is initiated unconsciously in the brain. In light of that, Libet called this essential function of awareness *conscious veto*.

At this point, however, Libet faced a conundrum: does the conscious veto have an unconscious origin? In other words, do we become aware of the choice to veto as we become aware of the wish to act? In this case, Libet's answer was firmly negative. According to him, the choice of veto would be an unconscious causal event, and, as a result, consciousness would be absolutely powerless. Libet based the solution of the problem on a subtle distinction between awareness and the contents of awareness. He admitted that full awareness of the decision to veto an action might require preceding unconscious processes, but he contended that the contents of that very awareness would have the same requirement. Thus, the process of awareness and its contents are to be thought of as two separate features of consciousness, so that the neurophysiological time delay that characterizes the former does not affect the latter.

According to Libet, the essential nature of the conscious veto is to exert a control function on the action planning, and this control can exclusively occur in the very moment of one's consciousness of the wish to move. Therefore, unconscious brain processes do not necessarily underlie the control function of conscious contents. Libet's idea, however, is based on a theoretical assumption that should itself find empirical evidence, that is, that the control function of the conscious veto can be associated with the activation of the same brain areas that are thought to be associated with the simple awareness of the wish to act. If this were true, there would not be a further delay caused by the activation of other cerebral pathways involving the conscious processing of the control function. But this picture might imply a fatal drawback for Libet's theory. In fact, if the control function does not need a further elaboration of different brain areas in addition to those which are involved in becoming conscious of the action plan, then the simple fact of being conscious of a wish to move will be *sufficient* to enable the control function. However, this does not seem to be the case in the real world, where there are movements which people are aware of but cannot stop. For instance, the involuntary tics performed by patients with Tourette syndrome fall into such a category. Libet refers to these tics as an example of movements performed in absence of free will. Yet, if we have to take for granted his concept of the conscious veto, then this kind of movements should occur not only in absence of free will but also *unconsciously* of any wish or urge to act. Since this is not true, it is apparent that the control function requires the engagement of other brain areas in addition to those which are

associated with the awareness of the will, whose activation would necessarily entail a further delay.

Libet's experiment has been replicated several times with the same results: a delay of consciousness with respect to the preparation of movement. Libet has thus been considered a champion of determinism by those who endorse this philosophical doctrine. He never thought of himself as such, however. In fact, he claimed that both determinism and indeterminism are unproven speculative beliefs. Therefore, Libet suggests to adopt the putative view that free will is real, based on his findings of a deliberate control on our actions and in consideration of the deep sense of freedom which we stubbornly experience.

Libet's experiment has been criticized in different ways with regard to both its construction and results. Criticism has mainly concerned the concept of readiness potential as the very moment in which the action is initiated, the reports of the subjects which may not be sufficiently reliable, and the possibility to precisely identify the time for the conscious decision. A thorough review of these critiques is beyond the scope of this chapter, but it can be noteworthy to put forward the following consideration. Libet's finding that consciousness is characterized by a degree of temporal delay seems to be sound. In fact, conscious processing needs to elaborate information in a more refined way compared to unconscious processing. In other words, consciousness appears to be more extended in time insofar as it implies the activation of widespread brain networks. On the other hand, the volitional process should not be easily fragmented in separate episodes, leading to outcomes derived from conscious pathways as opposed to outcomes derived from unconscious processes. Indeed, volition should be conceived as a continuous flow of both conscious and unconscious processes that appear to be merged and indivisible.

In addition to his controversial research on free will, Libet is also the proponent of an original theory on the nature of consciousness. According to him, the unitary and integrated flavor of conscious experience is the essential feature that a reliable theory of consciousness should be able to explain. In particular, science should address the profound question of how particularized and multifarious neural patterns can give rise to the coherent and unified phenomenology of consciousness. To solve this problem, Libet introduces the original concept of *conscious mental field* (CMF). This field would be different from any other physical field (electromagnetic, gravitational, etc.), since it could not be described in terms of observable physical constituents, being accessible only by the subject of the conscious experience. Therefore, the CMF would fall in a peculiar phenomenological category and, by functioning as mediator between the nonphysical mind and neurophysiology, would account for the emergence of subjective experience. Moreover, such a field would allow communication in the brain without the help of neural pathways, thereby ensuring a causal role for consciousness.

Libet bases the verifiability of his theory on this communicative property. He argues that it would be possible to isolate in situ by means of a surgical technique a living slice of brain from all the other cerebral areas. Thence, the CMF theory would be empirically verified if the contents of conscious experience of this isolated part continued to influence the behavior of the subject. This operation may be carried out

on patients who would benefit from a resection of part of the brain, although ethical constraints have prevented the realization of such an experiment.

Libet claims that the CMF is a nonphysical and nonreducible emergent property of the brain. But at the same time, he holds that CMF could not exist without cerebral activity. Thus, although Libet declares himself a non-Cartesian dualist, his position could be considered an original reappraisal of Cartesian dualism, with the difference that for Descartes mind is to be thought of as a separate nonphysical entity independent of matter, while for Libet mind is to be thought of as a separate nonphysical entity dependent on matter. Still, the mechanism by which the *physical* activity occurring in the brain creates an emergent *nonphysical* property like the conscious mental field (which, in turn, can have a reentrant effect on cerebral processes) remains unclear.

Essential Bibliography

- Libet B (1999) Do we have free will? *J Conscious Stud* 6:47–57.
This article presents a detailed account of Libet’s famous and groundbreaking experiment on free will. The ethical implications of the veto hypothesis are also discussed, with stimulating theoretical considerations with regard to the philosophical theories of determinism and indeterminism.
- Libet B (2004) *Mind time: the temporal factor in consciousness*. Harvard University Press, Cambridge, MA.
This book encompasses all the pioneering researches led by Libet on consciousness and free will. The author proposes an intriguing reappraisal of Cartesian dualism and put forward a scientific and testable hypothesis as to the nature of consciousness.

The Sensorimotor Theory of Consciousness

...we can simply say that an organism consciously feels something when it has at least a rudimentary self which has access consciousness to the fact that it is engaged in a particular type of sensorimotor interaction with the world.
(Why Red Doesn't Sound Like a Bell)

John Kevin O'Regan (born on 24 May 1948) is an American neuroscientist who has lived and worked in Paris since 1971. He studied theoretical physics at Sussex and Cambridge Universities and in 1975 gained a PhD from Cambridge University for his research on eye movements during reading tasks. His interest in the study of the perceived stability in the visual process has led him to the discoveries of an *optimal viewing position* for the eye to fixate in words and the phenomenon of *change blindness*. In change blindness – codiscovered with collaborators Ron Rensink and Jim Clark – a person looking at a picture is not able to see major changes that occur in a given scene when it is presented again with changes accompanied by a brief interruption like a cinema cut, a blank, or even small distractors like mudsplashes on a car windscreen. Kevin O'Regan has significantly contributed to develop the so-called *sensorimotor approach* to sensation and consciousness, according to which both sensation and consciousness involve an interaction with the body and the environment. In particular, O'Regan has addressed the problem of the nature of phenomenal consciousness – the *what it is like* of being in a certain sensation rather than another (see Chaps. 1 and 11) – both theoretically in relation to space and color perception and experimentally in relation to pain and sensory substitution (i.e., the possibility of using one sensory modality, e.g., hearing, to replace another, e.g., vision). O'Regan is also interested in applying his work to robotics. He is the current Director of the Laboratoire Psychologie de la Perception, CNRS, Université Paris Descartes.

O'Regan starts his analysis by asking what a feel or phenomenal conscious state really is. He identifies some extra components that accompany a feel. First, there are the *cognitive states* that a feel can evoke. For instance, seeing red could evoke images of roses, traffic lights, blood, and ketchup. These are all mental associations that do not constitute what it is like to experience red but can be added to the feel of seeing red. Second, there are the *learned bodily reactions* that the feel can cause, which derive from the habits that have been associated with the experience of that certain feel, for instance, pressing on the car brake when one sees that the traffic

signal is red. These bodily reactions are similar to the mental associations: they are not related to the feel but can be add-ons to what happens when someone has that sort of experience. Finally, other aspects of the feel are the *physiological states or tendencies* that the feel generates. In the case of seeing red, specific physiological manifestations can occur in the body, involving changes in heartbeat as well as in the autonomic nervous system, for example, when someone faints after having seeing a bleeding wound. Similarly to the cognitive states and learned bodily reactions associated with the feel, these physiological effects too are not essential in the constitution of the actual *raw feel* of redness.

With the expression “raw feel,” O'Regan designates the conscious experience in itself, which is what remains after all the add-ons have been stripped away. It is as if the experience is peeled to its core, just like a fruit is peeled to its fleshy pulp (Fig. 26.1).

A raw feel is, essentially, the *what it is like* to have the experience; in other words, the particular flavor or basic feature that constitutes the experience in itself and is called by philosophers *quale* (about the philosophical concept of *quale* or *qualia*, plural of *quale*, see Chaps. 1, 5, and 13). The explanation of raw feels or *qualia* is thought to be the difficult or hard problem of consciousness, and, according to O'Regan, a theory of the conscious mind will be comprehensive only if it is able to account for the nature of raw feels. O'Regan claims that there is no way to tackle this problem satisfactorily if we conceive of the raw feel as being wholly generated in the brain. In order to better explain this point, he invites us to suppose that advances in neurosciences make it possible to localize the brain networks that can provide the raw feels of red and green, respectively. He also asks us to imagine that two groups of neurons have been detected and characterized by different connections and firing rates. Thus, we could find that one specific pattern and frequency oscillations of neurons are responsible for the feel of red and another pattern and frequency oscillations of different neurons are responsible for the feel of green. However, O'Regan points out that these mechanisms do not provide an answer to the problem of raw feel, because we can always ask why this particular neuronal

Fig. 26.1 Willem Klaf (1619–1693), *Still Life with a Nautilus Cup* (detail), oil on canvas (1662), Thyssen-Bornemisza Museum, Madrid, Spain (Image cropped by authors)



pattern brings about the feel of red rather than the feel of green. Why certain firing rates and neurotransmitters are specific to green and others to red? Why could it not be the other way round?

According to O'Regan, finding the brain structures that generate a particular raw feel or phenomenal conscious state helps explain the *extra* components of feel but not what we would really like to know about raw feels or phenomenal conscious states, namely, why they are the way they are. In fact it could seem that the brain and the conscious mind are incommensurable with each other, as there are no precise links that can lead from certain neuronal patterns to certain raw feels and vice versa. However, from the fact that we cannot explain why particular brain processes generate the conscious sensation of red, it does not follow that these particular brain processes are not sufficient and necessary in order to consciously see red. Contrary to O'Regan's analysis, philosophers and neuroscientists who endorse the materialist identity theory of mind could reasonably object that asking what are the links between brain processes and raw feels is an ill-defined question, because there are no such links, as brain processes and raw feels are exactly the same things (for this materialist objection, see Chap. 2). Thus, the identity between brain processes and raw feels could eventually be just a fact about nature.

O'Regan identifies four "mysteries" about raw feels that his sensorimotor theory is able to clarify. First, having a raw feel implies feeling like something: any raw feel is in fact characterized by a sort of sensory presence or phenomenal quality. Second, there is a wealth of phenomenal qualities that can characterize raw feels, such as the heat of fire, the softness of cotton, the green of leaves, the sound of cello, and so on. Third, there seems to be a sort of structure among the many differences across raw feels, so that it is possible to make comparisons between them in terms of dimension types. Two examples of dimensions commonly used to describe raw feels are intensity and length of time. Fourth, even though it can be somehow possible to compare raw feels, their basic and intrinsic nature (the *what it is like* to have them) remains ineffable; it cannot, in other words, be described in any way. This is to say, someone will be able to know the smell and taste of the white truffle only if he or she smells and tastes it.

The core of the sensorimotor approach is that the feel does not strictly occur in the brain; rather, it resides in noting that the body is interacting in a particular way with the environment. The feel is therefore to be conceived of as *a mode of interaction with the world*. Following this line of reasoning, O'Regan claims that it is possible to solve the four mysteries of raw feel by analyzing how the sensorimotor interaction takes place. In order to build a sensorimotor framework, within which every mode of bodily interaction with the world can be accounted for, O'Regan pinpoints four concepts: richness, bodiliness, insubordinateness, and grabbiness.

First, feels resulting from interaction with the world are much richer in details than memories and imagines. It is also impossible to create a new imagine that is not related to any known one. In a sense, the contents of imagination are parasitic on the world. Second, feels have bodiliness: whenever the body is moved, there is an immediate change in the incoming sensory input. Think, for instance, of touching a surface with the hand or sniffing the air by moving the head so as to identify the place which

the smell comes from. On the contrary, remembering and imagining are unaffected by bodily movements. Third, feels are also insubordinate, because the sensory input is not totally controlled by the body and can change independently of bodily motion. In fact, no one can make a feel appear and disappear. Fourth, feels have the feature of grabbiness: the sensory system is wired up with an alerting mechanism that can grab cognitive resources in certain circumstances. In hearing, for example, a sudden noise can entirely grab someone's attention toward the source of the sound.

By making use of these four characterizing properties, O'Regan's sensorimotor theory suggests a way to explain the mysterious aspects of phenomenal conscious experiences or raw feels. Raw feels have distinctive phenomenal qualities (they feel like something rather than nothing) because they possess certain degrees of richness, bodiliness, insubordinateness, and grabbiness. Differences and analogies between raw feels are then explained by the fact that the body interacts in different ways with the world but always on the basis of shared sensorimotor laws and constraints. Ineffability, on the other hand, remains inescapably unaccountable; however, the sensorimotor approach has at least the advantage of providing descriptions in terms of dimensions (richness, bodiliness, insubordinateness, and grabbiness) that are commensurable to ordinary language.

In addition to the four properties of the body-world interaction, the concept of *self* is another key aspect of O'Regan's sensorimotor theory. O'Regan distinguishes between the cognitive and social dimensions of the self. Within his sensorimotor approach, the cognitive dimension plays the most important role. Following the classification proposed by Bekoff and Sherman (2004), O'Regan identifies a continuum of three categories within the cognitive self: *self-distinguishing*, *self-knowledge*, and *knowledge of self-knowledge*. At the most basic level, the cognitive self can act differently with regard to parts of its own body according to the stimuli coming from the outside world. Self-distinguishing does not require either a brain or a mind; an example is the immune system, which can distinguish foreign cells from body's own cells. Instead, self-knowledge requires a system of cognitive capacities, which is able to choose plans of actions and make judgments. At this level, the body is recognized as being a separate entity from its environment, but the mind may still not know that it possesses this kind of knowledge. In order for a mind to recognize itself as being an individual among other individuals, it needs to achieve a meta-knowledge, that is, it has to know that it has self-knowledge. At this stage, an individual can develop a theory of mind by considering other individuals as having beliefs, intentions, desires, and goals, in analogy to the beliefs, intentions, desires, and goals that the individual itself has.

According to O'Regan's sensorimotor approach, the concept of self is strictly associated with that of conscious experience or raw feel. In fact, the latter becomes possible only on the basis of the former. Borrowing the notion of *access consciousness* proposed by philosopher Ned Block (1996) and interpreting it as a phenomenon in which an agent has cognitive access to the fact that it has cognitive access to something, O'Regan claims to have all the ingredients that are necessary to explain phenomenal consciousness. These include a sensorimotor interaction with the world and a self able to experience and having cognitive access to the interaction. Thus,

given these three conditions, the particular *what it is like* of raw feel results from the cognitive access that a self-agent has to having a certain sensorimotor interaction which involves the aspects of richness, bodiliness, insubordinateness, and grabbiness. The degree of consciousness will vary according to the degree of cognitive access the self-agent has to the quality of its experience. Therefore, full-blown phenomenal consciousness can be achieved only by the self-agent with the highest degree of cognitive access (knowledge of self-knowledge).

The sensorimotor theory has the merit to broaden the approach to the study of consciousness, as well as providing also interesting suggestions to the science of robotics. It is however debatable to claim that the solution to the problem of phenomenal consciousness will be found by giving more emphasis to world-body interaction rather than to brain activity. Arguably, a sensorimotor approach can be of great value in explaining how the contents of consciousness are generated. However, it is not clear whether this theory might give a comprehensive account of consciousness as a neurobiological property (see also Chap. 12 for further discussion on this point), since a panoply of neuroscientific findings suggest that the conscious mind is deeply rooted in brain processes.

Essential Bibliography

- O'Regan JK (2011) *Why red doesn't sound like a bell: understanding the feel of consciousness*. Oxford University Press, Oxford.
The book is a thorough exposition of the sensorimotor approach to conscious experience. The author deals with the problem of qualia or raw feels and identifies four mysteries about these phenomena. In the attempt to provide a convincing account of these mysteries, he develops a sensorimotor framework based on the features of richness, bodiliness, insubordinateness, and grabbiness.
- O'Regan JK (2012) How to build a robot that is conscious and feels. *Minds and Machines* 22:117–136.
Building on the sensorimotor arguments put forward in *Why Red Doesn't Sound Like a Bell*, the author develops a more pragmatic point of view about the concepts of “phenomenal consciousness” and “feel,” so as to investigate to what extent these processes might be implemented in nonbiological machines.

Consciousness and Quantum Physics

In our own brains, the OR [objective reduction] process that evoke[s] consciousness, would be actions that connect brain biology (quantum computations in microtubules) with the fine scale structure of space-time geometry, the most basic level of the universe, where tiny quantum space-time displacements are taken to be responsible for OR.

(Consciousness in the Universe, A Review of the 'Orch OR' Theory)

Sir Roger Penrose (born on 8 August 1931) is a British mathematician who has given outstanding contributions in mathematical physics, especially within the fields of general relativity and cosmology. He has also collaborated with British theoretical physicist and cosmologist Stephen Hawking, with whom he shared the Wolf Prize for physics in 1988. Penrose studied at University College London, where he graduated in mathematics. In 1958, he earned his PhD from the University of Cambridge. He currently is Emeritus Rouse Ball Professor of Mathematics at the Mathematical Institute of the University of Oxford and Emeritus Fellow of Wadham College.

Stuart Hameroff (born on 16 July 1947) is an American anesthesiologist who made important contributions to consciousness studies. He gained his BS degree from the University of Pittsburgh and his MD degree from Hahnemann University Hospital. In 1975, he joined the University of Arizona, where he became professor in the Department of Anesthesiology and Psychology and Associate Director for the Center for Consciousness Studies (1999). In 2003, he became Emeritus Professor for Anesthesiology and Psychology at the University of Arizona.

Penrose and Hameroff have developed a thought-provoking theory about the nature and origin of consciousness. Their idea is that we need a quantum mechanical approach in order to account for the extreme complexity of human brain conscious functioning. Quantum physics is the branch of mechanics that deals with the mathematical description of the motion and interaction of subatomic particles. This theory is one of the greatest achievements of the human mind and incorporates the concepts of quantization of energy, the wave-particle duality, and the uncertainty principle. The hypothesis of quantization states that at the smallest length scales, reality can be described as composed of discrete packets of energy (quanta). In other words, each physical property can be quantized in order to assume only certain values. With regard to light, for instance, a photon is a single discrete quantum of

light. The principle of the wave-particle duality states that every quantum of energy can exhibit the properties of both a particle and a wavelike phenomenon. Finally, the uncertainty principle states that the momentum (i.e., the quantity of motion of a body, measured as a product of its mass and velocity) and position of a particle cannot be precisely determined at the same time. If the momentum is known, then the position remains uncertain and vice versa.

Penrose and Hameroff identify three theoretical possibilities which can broadly encompass the majority of options proposed by thinkers in order to understand conscious events. The first possibility considers consciousness as a property dependent on conventional physical processes, a natural evolutionary product of the biological development of the nervous system. In light of this view – which describes the prevalent standpoint in current neuroscience – the conscious mind originates from complex mechanisms of biological computation. Opinions differ as to when, where, and how consciousness appeared in natural history, for instance, whether it appeared only recently in humans or earlier in lower organisms. Within this framework, consciousness can be thought of as an epiphenomenon, a by-product of the brain which is not able to causally influence behavior (see Chap. 22) or, by contrast, as playing an effective role in causing human actions (see Chaps. 15, 19, and 28). In either case, however, consciousness does not appear to be an intrinsic feature of nature.

The second theoretical possibility is to consider consciousness as a separate entity with its distinctive qualities, which cannot be reduced to physical actions and cannot be controlled by physical laws. According to this view, consciousness has always been present in the universe, although its nonphysical nature makes it impossible to explain this phenomenon by science. Historically, this hypothesis has been exemplified by Descartes' position (see Chap. 6) and by idealism, namely, the philosophical doctrine contending that consciousness is all that exists, the material world being an illusion. The pervasive presence of consciousness in matter is also at the root of panpsychism, another philosophical doctrine which ascribes consciousness to entities independently of the combination of their material elements (see Chaps. 1 and 17 for contemporary authors who have shown inclination for panpsychism).

The third theoretical possibility – which is the one endorsed by Penrose and Hameroff – is to think about consciousness as the result of discrete physical events, which have always existed in nature as noncognitive and proto-conscious events. According to this view, living creatures evolved a complex mechanism able to *orchestrate* (that is to say, organize and isolate from random environmental interferences) these discrete physical events so as to couple them with specific brain activity. This coupling is what ordinarily happens within our brains and results in meaningful conscious states that can in turn cause and control behavior. Consciousness therefore appears to be an intrinsic feature of a specific quantum mechanical interaction in the texture of nature.

Penrose and Hameroff propose that the discrete physical events are specific moments of quantum computations within the brain. These events are characterized by experiential qualities and constitute an *orchestrated objective reduction* (Orch OR) of the superposition of states due to the quantum mechanical duality. Thus, consciousness consists of an organized sequence of distinct events occurring in the fine-scale structure of space-time geometry, which is the basic level of reality. The events of objective reduction are ubiquitous within physical interactions; within the brain, however, Penrose and Hameroff suggest that they specifically occur in tiny tubular structures located inside neurons, which are called *microtubules*. Microtubules are present in huge numbers in the cytoplasm of every cell (not only of neurons), sometimes aggregating to form more complex structures. They are thought to be an important component of the cytoskeleton, a scaffolding-like protein network which maintains the cellular structure. Microtubules constitute platforms for intracellular transport and are therefore involved in a variety of cellular processes, including the movement of vesicles, organelles, and intracellular substances.

According to Penrose and Hameroff's theory, microtubules are thought to be a sort of automata able to represent and process information via calculations based on tubulin (i.e., the protein that is the main constituent of microtubules) dipole states. As tubulin dipoles can combine in several different states, they have a high capacity for integrating information. This exceptional computational power relies on quantum mechanical laws, according to which quantum-superposed states can simultaneously proceed until an objective reduction or "choice" occurs. Penrose and Hameroff have provided calculations so as to predict when the state-reduction process might occur. What is more, the "choice" involved in the objective reduction process would be accompanied by a proto-element of experience, which Penrose and Hameroff refer to as *proto-consciousness*.

Penrose and Hameroff suggest that the most logical site for microtubule Orch OR (and, consequently, emergence of consciousness) would be in postsynaptic dendrites and soma (cellular body) of neurons. During the integration phases of the firing neuronal patterns, microtubules could be stabilized and orchestrated according to quantum dipoles. This orchestration would in turn lead to tubulin superposition and quantum integrated computation. As a result, consciousness would appear as a particular sequence of configurations of space-time geometry, configurations that are singled out of previous quantum states superpositions.

According to Penrose and Hameroff, a quantum mechanical approach to consciousness can account for all the manifestations of human behavior, including the features of the human mind that seem to exceed explanations provided in classical computational terms, such as the faculty of understanding, intuition, creativity, and, ultimately, free will. The brain would in fact exploit the property of quantum physical systems to be in multiple superimposed states in order to explore a number of different options in a small amount of time. In other words, before the Orch OR

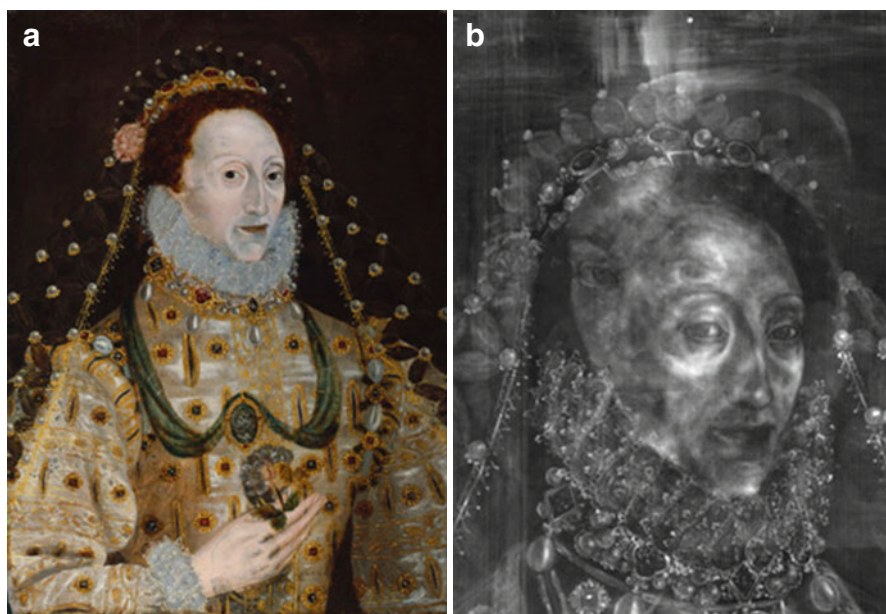


Fig. 27.1 X-ray imaging on the portrait of Queen Elizabeth I revealed that the canvas hid a first version subsequently deleted by the painter. (a) Unknown artist, Queen Elizabeth I, oil on panel (1580–90), National Portrait Gallery, London, UK. (b) Detail from the X-ray mosaic, showing the overlaid portraits

takes place, the brain would be able to elaborate more than one scenario at the same time, by superimposing different quantum states of the world, similarly to what we can captivantly see in the X-ray mosaic of the overlaid portraits of Queen Elizabeth I (Fig. 27.1).

Penrose and Hameroff's theory has not been immune from criticism by both physicists and neuroscientists. Some physicists have pointed out that the warm-blooded bath that surrounds the brain in the skull is incompatible with quantum computing, which requires extremely cold temperatures in order to isolate a system and prevent decoherence (i.e., the phenomenon that occurs when quantum superposition collapses into a definite state). In turn, neuroscientists have pointed out that the time scale of consciousness is different to the small scale (10^{-15} s) at which quantum decoherence typically occurs. Furthermore, although quantum phenomena might somehow influence brain activity, their intrinsic indeterminacy might not offer a satisfactory solution to the problem of free will, as, for instance, neuroscientist Stanislas Dehaene (see Chap. 19) pointed out in his recent book on consciousness. In fact, nothing would be more distant from our idea of free will (which refers to an autonomous decision-making process) than a mere form of quantum randomness occurring within the brain.

Essential Bibliography

- Penrose R (1989) *The emperor's new mind: concerning computers, minds, and the laws of physics*. Oxford University Press, Oxford.
The book intriguingly merges physical concepts and philosophical arguments, including the issue as to whether or not artificial intelligence will ever be able to achieve consciousness. In this work, Penrose argues that there are some aspects of human thought processing that cannot be reproduced by machines.
- Penrose R (1994) *Shadows of the mind: an approach to the missing science of consciousness*. Oxford University Press, Oxford.
Penrose further develops the line of reasoning put forward in his previous book by adding more arguments to the claim that there is something in the conscious mind that transcends classical computation. A quantum mechanical approach to consciousness is suggested to provide a solution to this problem, along with the hypothesis that neuronal microtubules could provide the interface where classical and quantum physics can meet.
- Hameroff S, Penrose R (2014) *Consciousness in the universe: a review of the 'Orch OR' theory*. *Phys Life Rev* 11:39–78.
This article provides a comprehensive (albeit at times technical) discussion of the orchestrated objective reduction (Orch OR) theory of consciousness. The hypothesis that quantum computations can occur within neuronal microtubules is thoroughly explained by Hameroff and Penrose, along with replies to its critics.

Consciousness as Integrated Information

*...whenever the mechanisms of a complex unfold
and specify informational relationships,
the flower of experience blooms.*

(Consciousness As Integrated Information: A Provisional Manifesto)

Giulio Tononi, born in Trento, Italy, is a psychiatrist and neuroscientist renowned not only for his theory of consciousness but also for his studies on sleep mechanisms. He is Professor of Psychiatry at the University of Wisconsin-Madison and has held faculty positions in Pisa, New York, and San Diego. He collaborated with Nobel Laureate Gerald Edelman, with whom he developed a theory on how the brain can integrate a critical amount of information in order to make it conscious. This theory has been further refined by Tononi and, at present, is one of the most promising scientific theoretical frameworks for consciousness research.

Tononi's approach identifies two issues that a theory of consciousness must address to account for how brain activity can generate conscious experience. The first issue is to define the conditions that establish to what extent a system has consciousness. To solve this problem, which concerns the quantity or level of consciousness, we need to identify what parts of the brain are important for generating consciousness and why these specific parts are important and others are not. We need to know, in other words, how a physical system can bring about and maintain conscious states. The second issue is to define the conditions that establish what kind of consciousness a system can have. By solving this second problem, which concerns the quality or content of consciousness, we would understand why specific conscious experiences (hearing the chimes of a clock tower, for instance) "feel" the way they do and why they differ.

The fundamental tenet of Tononi's model – the *integrated information theory* – claims that consciousness is a specific process by which information is integrated. Tononi adopts a functional definition of information, according to which information is the cutback of uncertainty among different outcomes when one of these occurs. Thus, a system which is able to discriminate among several possible states will have a higher degree of differentiation with respect to another system that is able to discriminate between two possible states only (say, "light on" and "light off"). In other words, the number of distinguishable states that a system can adopt corresponds to the amount of information that can be encoded by the system.

According to the information integration theory, however, a system that is only able to distinguish among possible states is not as yet conscious. In fact, in order to be conscious, a system must also be able to integrate information in a spatiotemporal framework, so as to construct a coherent and unitary picture of reality, which cannot be further reduced to elementary components. From the phenomenological point of view, the process of integration is evident in every conscious experience. We cannot independently experience parts of an image but always the image as an integrated whole. For instance, colors are not independent of shapes, shapes are not independent of objects, and objects are not independent of their localization. Likewise, the spatiotemporal scale of consciousness is always the same: we cannot voluntarily speed up or slow down our perceptions.

Thus, information integration theory claims that a system which is capable of both discriminating among multiple possible states of the world and integrating information is necessarily conscious. Furthermore, the degree of discrimination and integration that a system is able to reach can be expressed in mathematical terms. Tononi indicates the value of information integration with the Greek letter Φ (phi, which is pronounced *fi*). Hence, the higher the value of Φ in a system, the higher the level of consciousness that the system can have. The unit that effectively integrates information is called *complex* by Tononi. Specifically, a complex is a set of elements that form a whole which has $\Phi > 0$ and is not included in a larger set of elements having higher Φ . Complexes are therefore the places where information is integrated and correspond to the real “subjects” of conscious experience. A complex, then, appears to have a specific and private perspective or point of view, as information can be integrated only within its boundaries. This is why consciousness seems to have an irreducible subjective nature. In turn, a number of complexes can combine to create a system in which the one with the maximum value of Φ is called the *main complex*. In addition to these properties, complexes can overlap, as the same element can be part of more than one complex, and at the same time be connected with elements that do not strictly belong to them and through which information is exchanged (*ports-in* and *ports-out*, respectively). Consequently, the set of elements that make up a complex and underlie consciousness are not static but can be considered to develop a *dynamic complex* (a concept that is equivalent to Edelman’s *dynamic core*, which we have discussed in Chap. 22).

By introducing his theory’s requirements in computer simulations, Tononi showed that a system with high Φ has networks that interlink specialized modules with other parts, which in turn are able to functionally integrate information coming from these modules. Tononi emphasizes that this network organization is very similar to the one we find in the mammalian brain, where highly specialized cortical areas can copiously interact through a large network of connections. In particular, scientific evidence suggests that the corticothalamic system can be a good candidate for the neural substrate that can integrate information, since an impairment of this system necessarily causes either alteration or loss of consciousness.

With his information integration theory, Tononi is able to answer both questions about the quantity and quality of consciousness. In fact, the quantity of consciousness, generated by a complex of elements, is determined by the amount of integrated

Fig. 28.1 M. C. Escher (1898–1972), *Order and Chaos* (1950). (M.C. Escher’s “Order and Chaos” © 2014 The M.C. Escher Company-The Netherlands. All rights reserved. www.mcescher.com)



information that the complex can elaborate, whereas the quality of consciousness is determined by all the informational relationships within the complex. Thus, the way the complex can generate integrated information determines both the degree and the content of consciousness. In addition, all the internal connections among the elements of the system develop into a topographic space, called by Tononi *qualia space*, in which every possible state of the system constitutes a dimension. Within this space, a specific set of connections generates a characteristic shape or *quale*, that is, the conscious feeling of being in a certain state. In a sense, the *quale* is like the “form” of a multidimensional crystal or, to use a more technical term, a polytope. In order to try to imagine a *quale*, we could use an artistic metaphor by referring to M. C. Escher’s lithograph *Order and Chaos* (Fig. 28.1).

In the center of Escher’s lithograph, we see a geometric solid surrounded by different shattered objects. This crystal-shaped figure might be considered like a *quale* generated by the brain: a highly ordered pattern of integrated information, which is produced by a complex. Every specific *quale* can therefore be the reflection of an external object. The analogy between Escher’s solid and a polytope is obviously simplified, but it helps to highlight a fundamental aspect of information integration theory, which is the capacity of complexes to derive order and meaning from a chaotic reality. The significance of a possible state is associated with the increase of the level of consciousness and depends on the number of discriminations that the system is able to do. This is a central point in information integration theory, because it implies that the greater the number of possible states in which the system can be, the more significant for the system to be in a specific state. In a sense, meaning appears to emerge naturally from the repertoire of possible states that a system is

able to generate. Consequently, the value of being in a specific state is strictly associated with the number of alternatives among which the system can choose in order to experience the world. A human brain, therefore, can have a much more refined and assorted experience of the world compared with a photodiode, which can only experience reality as “light on” or “light off.”

It follows that, according to the information integration theory, an experience is a particular shape in the *qualia space* and that this shape or *quale* univocally and entirely characterizes the quality of the conscious content. Each experience is specified by a determinate shape and not by others, so that among individuals conscious experiences can be similar or different, to the extent that their shapes are similar or different. This picture has important philosophical consequences. In fact, philosophical zombies or creatures that are exactly the copy of another person without being conscious are not conceivable within the framework of information integration theory. We have seen in Chap. 1 that this is one of the main arguments used by philosophers to discredit identity theories of mind. However, if consciousness is integrated information, it must follow by definition that each system with the capacity to integrate information, so as to create a space of *qualia*, is also conscious.

The kind of identity defended by the model of integrated information is not identity between conscious experiences and brain processes but between conscious experiences and shapes of informational relationships. In fact, according to Tononi’s theory, not only human brains but also other physical systems could produce the same conscious experience, provided that they are able to construct the shape of informational relationships which is distinctive for that particular experience. Thus, information integration theory has a straightforward position on the issue of whether AI can be conscious: an artificial machine will necessarily be conscious, insofar as it is capable of generating a critical amount of integrated information. This stance is in striking contrast with the claim that both meaning and subjective experience are exclusive properties of biological matter. As we have said in Chap. 15, when we discussed the “Chinese room” mental experiment by philosopher John Searle, only empirical advancements in science will be able to address this issue definitely.

With regard to consciousness, other two implications of information integration theory are worth mentioning. The first implication is that consciousness does not appear to be an all-or-none phenomenon. Since consciousness is an intrinsic property of integrated information, any system that can elaborate information in this way has some degree of conscious experience. Therefore, the model of integrated information raises the intriguing question as to what level of Φ becomes relevant in order to have a full-blown consciousness. In fact, if the model is correct, it would be possible to measure with precision the level of consciousness of any system, independent of its ability to account for it. The second important implication is that consciousness does not seem to require many of the features that we naturally associate with being human, such as emotions, memory, language, self-reflection, and sense of agency. A system should not need these faculties to integrate information. This is a particularly bad news for the theoretical approaches that consider the contribution of emotions as fundamental for developing consciousness, such as those discussed in Chaps. 18 and 23. However, the assumption that emotions,

memory, and other typical human faculties (such as intentionality) play no role in the generation and maintenance of consciousness appears to be rather counterintuitive. In light of that, the model of integrated information would neither explain nor measure consciousness but, on the contrary, the most essential property of computational systems: their computational power. And this property would be necessary, but not sufficient for consciousness.

Only future will tell if Tononi's theory will stand the test of time. Anyway, as information integration theory requires that consciousness can be completely expressed and explained in pure mathematical terms, Galileo Galilei's suggestion that mathematics is the fundamental language of the world seems near to come true. And if neuroscience can prove it, we should be expected to radically change our concept of consciousness, as well as the way we see ourselves as human beings.

Essential Bibliography

- Tononi G (2012) *Phi: a voyage from the brain to the soul*. Pantheon Books, New York.

This book tells the story of how the seventeenth-century Italian mathematician and astronomer Galileo Galilei, one of the fathers of modern physics, discovers the marvels of the brain. The author imagines that Galileo, with the help of three learned guides that resemble Sir Francis Crick, Alan Turing, and Charles Darwin, undertakes a fantastic voyage in order to understand the mysterious nature of the relationship between the brain, mind, and matter.

- Tononi G (2004) An information integration theory of consciousness. *BMC Neurosci* 5:42.

In this article, Tononi described in great detail the information integration theory of consciousness. The reader is first introduced to the innovative concepts of Φ , *complex* and *qualia space*, and then led to the "astonishing hypothesis" that a shape of interconnections in the *qualia space* corresponds to the quality of conscious experience.

- Tononi G (2011) Consciousness as integrated information: a provisional manifesto. *Biol Bull* 215:216–242.

This paper puts the information integration theory within a global perspective, with a thorough discussion of its unexpected implications and consequences. Through a painstaking analysis, Tononi puts forward thought-provoking suggestions on the nature and development of consciousness.

Reflexive Monism

*...each human participates in a process
whereby the universe differentiates into parts
and becomes conscious in manifold ways of itself,
making the entire process reflexive.*
(Reflexive Monism)

Max Velmans, born 27 May 1942, is Emeritus Professor of Psychology at Goldsmiths, University of London. In 1994, he cofounded the *Consciousness and Experiential Psychology Section* of the British Psychological Society and made several publications in the area of consciousness studies, which culminated in a book titled *Understanding Consciousness*. In this book – which thus far has had two editions, in 2000 and 2009 – Velmans put forward his theory of consciousness (called *reflexive monism*) by trying to reconcile the phenomenal aspects of the mind with the physical aspects of the brain.

Reflexive monism is presented by Velmans as an alternative to both dualism and reductionism. It opposes dualist accounts of consciousness because it considers the world simply composed of one fundamental element. It also opposes the reductionist paradigms of both eliminativism and functionalism, because it discards the possibility of explaining the phenomenal mind in physical terms. According to Velmans, if we take for granted a monist perspective, there are three possible ways in which mental and physical features of the world can relate to each other:

1. The mind may be a particular aspect or arrangement of physical matter (as it is assumed by physicalist and functionalist theories of the mind).
2. The physical matter may be a particular aspect or arrangement of the mind (as it is assumed by idealism).
3. Mind and physical matter may be aspects or arrangements of something different, that is, a more elementary ingredient of reality which is in itself neither mental nor physical (as it is assumed by neutral monism and dual-aspect theory).

Reflexive monism is a nonreductionist dual-aspect theory of mind: as such, it lies within the third hypothetical framework. In this view, the material stuff of the brain and the mind are *one and the same thing, however observed from different angles*. Echoing philosopher Arthur Schopenhauer, who said that “motives are causes experienced from within,” Velmans could say that mental states are brain

processes experienced from within, just like neuronal patterns are mental events observed from outside.

The first important implication of this model is that the mind cannot be reduced to the brain, because mental states are located within the same ontological level of brain physiology. In other words, neuronal patterns are no more fundamental than mental processes. The second important implication of reflexive monism is that the physical and phenomenal worlds are not to be considered as divided fields of reality. We should therefore abandon our dualist common way of thinking, according to which the internal content of a conscious experience is a distinct representation of an external different object. In other words, we should become used to think that the conscious and phenomenal mind cannot be neatly separated from the physical world. In fact, just as the boundaries of physical processes are not precise, so the boundaries of phenomenal processes should be imprecise as well.

Velmans calls *phenomenological internalism* the common dualist view which locates phenomenal experiences either nowhere in the physical space (as it is held by Cartesian dualism) or within the boundaries of the brain (as it is held by both representational and reductionist theories of consciousness). Velmans argues that the dualist approach systematically discounts the importance of subjective experience. In fact, from the first-person point of view, the content of a perception is neither “nowhere” nor “inside the brain,” but it seems to be simply located in front of the perceiver, out there in the world. Thus, Velmans claims that the first-person account of experience provides us with phenomenal evidence which cannot be discarded: conscious appearances are really where they seem to be, namely, in the world in case of a perceived object, in the foot in case of a pain resulting from stepping on a thorn, etc. Consequently, there is no need for a perceived object to be accompanied by an additional representation of the object in the brain nor is there need for the pain in the foot to be accompanied by an additional experience of pain in the brain. In light of this, percepts or phenomenal representations in the brain are but theoretical fictions.

By contrast, Velmans’ reflexive theory of consciousness seems to reject the common dualist view by adhering to a form of *phenomenological externalism*, according to which the content of a perception (say, a cat) and the conscious experience of it are one and the same thing. In a sense, there is no real percept in the observer’s brain that produces or is subjectively experienced as a cat. All what there is, instead, is a complex brain activity that perceptually projects the phenomenal appearance of a cat in an external phenomenal space.

The concepts of *perceptual projection* and *phenomenal space* play a crucial role within this framework. Perceptual projection is a specific psychological effect which derives from unconscious perceptual processing and gives the reflexive flavor to Velmans’ model. Velmans highlights that this mechanism refers to observable effects which are empirically evident (namely, that the perceived cat is exactly where it seems to be – for instance, on the table – rather than in a representational construction inside the observer’s brain). In turn, phenomenal space is the place where all the projections or phenomenal appearances are located. Importantly, this is the actual space which we commonly experience ourselves in and the space

through which we move. In the first instance, it is the space that can be measured by our body and in which objects and events appear to be arranged with respect to us in specific ways (e.g., closer, behind, to the left, smaller, etc.).

Interestingly, reflexive monism is not strictly externalist with regard to all experiences. Cognitive processes that result in thoughts (such as memories) may well be located in the head or in the brain, as they are experienced to have an internal origin. In fact, Velmans claims that the location of an experience is an empirical matter, which entirely depends on its phenomenology. Consequently, the last word on where experiences really are is left to a systematic phenomenological analysis based on the subject's perspective, which can provide the evidence that the phenomenal and physical realities are substantially the same thing.

Thus, reflexive monism considers the first-person perspective phenomenological access to the world essential for understanding consciousness. Overall, Velmans' theory remains a realist doctrine, that is, a doctrine that believes things and objects are really out there, in the external world. What is more, reflexive monism claims a perfect overlapping between the phenomenal and physical reality. In a sense, any perceivable thing can be considered as having a sort of "phenomenal ghost," which is located in the exact place where the real thing is located. The only difference is that the "real thing" has an observer-independent location (which is relative to the other things located in the real space) and continues to exist even if the observer does not perceive it, whereas its "phenomenal ghost" has an observer-dependent location (which is relative to the other appearances in the phenomenal space) and ceases to exist as soon as the observer no longer perceives it. An important consequence of this intriguing picture is that whenever we try to determine the location of an object among other objects we are actually trying to measure the location of an *experience among other experiences*. This is one of the most significant implications that reflexive monism suggests and, according to Velmans, gives further support to the claim that conscious experiences cannot be measured and located in the brain.

In sum, Velmans' reflexive monism is to be regarded as monist from the ontological point of view, because it holds that only one kind of reality with both mental and physical properties exists in the world. However, this theory can also be considered as a form of dualism from the epistemological point of view, because it claims that the subjective knowledge given in the first-person perspective (what it feels to be in a specific mental state) cannot be reduced to the objective knowledge given in the third-person perspective.

Since the advent of novel neuroscientific investigating tools that made possible the study of the brain *in vivo*, the exact relationship between the first-person perspective and the third-person perspective has become one of the most debated topics of both neuroscience and philosophy of mind. Although the scientific paradigm requires objective and replicable data (third-person perspective), there is growing consensus that the phenomenological insights acquired from a subjective standpoint (first-person perspective) constitutes a fundamental aspect of any conscious experience and cannot be reduced to more elementary components. Thus, neuroscientists arguably need a common cognitive model that could simultaneously account for both subjective and objective features. Undoubtedly, reflexive monism is a pioneering attempt in this

direction. But in so doing, it is important to resist the temptation of objectifying the subjective standpoint. In fact, the use of general and uniform concepts such as *perceptual projection* carries the risk of objectifying the subjective perspective, which is, by definition, unique and distinctive for every individual and, thereby, cannot be subsumed under objective categories. As a result, the key concept of *perceptual projection* coined by Velmans would suffer the same fate of other objective categories described within the third-person vocabulary: arguably, the concept of *perceptual projection* could also be regarded as a theoretical fiction, as it has been claimed by Velmans with regard to the concept of phenomenal representation.

Reflexive monism considers consciousness as a fundamental and natural property of the universe. Consciousness emerges spontaneously by virtue of a reflexive mechanism wherein some of its components (human beings) have at the same time conscious understanding of the universe as a whole and of themselves as differentiated parts. As in the ancient myth of Narcissus, the cosmos can really contemplate itself indefinitely (Fig. 29.1).

This is so because the essential ingredient of nature has both mental and physical properties and therefore has the potential to manifest both physically and consciously. Importantly, each of these two properties cannot be explained in terms of the other one. This is why reflexive monism is thought to be a nonreductionist theory of mind. However, although reflexive monism is nonreductionist about the



Fig. 29.1 Michelangelo Merisi, known as Caravaggio (1571–1610), *Narcissus* (circa 1597–99), Galleria Nazionale d'Arte Antica, Palazzo Barberini, Rome, Italy (Image cropped by authors)

possibility to describe the mind in physical terms, it should be open to the possibility to reduce both mind and matter to the fundamental ingredient which composes all what there is out there in the world.

Essential Bibliography

- Velmans M (2009) *Understanding consciousness*. Routledge, London.
This book analyzes several philosophical approaches to consciousness and contends that, instead of starting from a theoretical perspective, the problem of consciousness should be tackled with the help of a detailed phenomenological approach to conscious experiences. Following this line of reasoning, the author puts forward his theory of reflexive monism as an alternative to both dualist and reductionist accounts of consciousness.
- Velmans M (2008) Reflexive monism. *J Conscious Stud* 15:5–50.
This article is a brief but extremely clear summary of Velmans' theory. Reflexive monism is compared to both the dualist and reductionist models, especially with regard to how a reflexive phenomenological approach can deal with the epistemological aspects of the mind-body problem.

The Theory of Multiple Consciousnesses

I believe that the search for the neural correlates of consciousness will be elusive until we acknowledge the many components of consciousness and their temporally hierarchical relationship to one another.
(The Disunity of Consciousness)

Semir Zeki (born on 8 November 1940) is one of the most distinguished contemporary neuroscientists. He studied at the University College London, where he gained a BSc in 1964 and a PhD in 1967, both on the subject of Anatomy. From 1980 to 1985, he was Henry Head Research Fellow of the Royal Society, and in 1981, he was appointed as Professor of Neurobiology at University College London, where he has become Professor of Neuroesthetics since 2008. His research work initially focused on the functional specialization within the visual brain. He then pioneered the study of the neurobiological basis of art and esthetics and contributed to establish a whole new field of neuroscientific studies called neuroesthetics.

With regard to the problem of the nature of consciousness, Semir Zeki has proposed that there are many consciousnesses distributed in time and space, even though we seem to deeply feel the unity of our conscious experience. Zeki claims that the property of the unity of consciousness has been overemphasized and the assumption that consciousness is a single unified entity could be a red herring in the quest for the neural correlate of consciousness (NCC; see Chap. 17 about the quest for the NCCs).

Zeki develops his theory of multiple consciousnesses by noticing the fact that we become conscious of different attributes of objects (such as location, color, orientation, and motion) at different times. This phenomenon is called *perceptual visual asynchrony*. For instance, the conscious perception of color, which is an attribute elaborated by a cortical area called V4, precedes by around 80 milliseconds the conscious perception of motion, which is an attribute elaborated by a separate cortical area called V5. The existence of perceptual asynchrony leads Zeki to suppose that we become conscious of color and motion in virtue of the activity of two brain systems which have distinct anatomical inputs. As a result, Zeki suggests that these two cortical areas are capable to produce two micro-consciousnesses, one for color and another one for motion, respectively.

According to Zeki, this hypothesis is further supported by neuropsychological studies, which show that lesions of V4 and V5 lead to different visual impairments: the former resulting in achromatopsia (acquired colored blindness) and the latter

resulting in akinetopsia (acquired visual motion blindness). Notably, a lesion in V4 does not affect the ability of V5 to process motion information, as well as a lesion in V5 does not affect the ability of V4 to process color information. Thus, a patient with achromatopsia is no longer able to perceive colors but can still perceive motion, whereas a patient with akinetopsia is no longer able to perceive motion but can still perceive colors.

Zeki therefore proposes that these sensory areas are the very places where consciousness begins. He claims that, at least at the level of the micro-consciousness, information does not need to be further elaborated in larger cortical networks. In fact, psychophysical-imaging experiments in humans provide evidence that the same specialized brain areas are active, regardless of whether the stimulus is consciously perceived or not. According to Zeki, the difference between a conscious state and an unconscious state rests in the fact that the neuronal activity of these brain areas is higher in the former state than in the latter. As a result, changes in the activity level of specific neuronal patterns could discriminate between conscious and unconscious processing, although it is not yet known the exact threshold beyond which consciousness can appear. Zeki also highlights that it is not yet clear whether this heightened activity of sensory regions is due to the recruitment of previously inactive neurons, to an increased discharge of neurons which were already active, or to an increase in synaptic input not accompanied by an increase in firing rate. Obviously, these hypotheses are not mutually exclusive.

Zeki suggests that the perceptual asynchrony implies a temporal hierarchy in the generation of consciousness. As the micro-consciousnesses originate in the activity of distinct brain sites at different times, it follows that these micro-consciousnesses are distributed in time and space, according to a hierarchical order. Zeki suggests that this fact raises an important issue about the binding problem, which is the problem as to how the brain is able to create a coherently unified perceptual scene from different perceptual stimuli. The issue concerns the way according to which the temporal hierarchy between the brain areas is organized. Does an area that has already elaborated its information “wait” for the others to finish their processing? If so, what kind of mechanism (e.g., a time buffer) is responsible for this waiting period?

The binding problem is one of the most difficult tasks that neuroscience has to face. A comprehensive theory of consciousness should be able to explain the unified and coherent scenarios that the brain creates by assembling disparate environmental stimuli. The common neuroscientific viewpoint is that the binding itself leads to the conscious experience. But Zeki’s hierarchical model of multiple consciousnesses logically implies that the binding has to be post-conscious. In other words, the binding between different attributes must occur only after these attributes have been elaborated by the brain specialized areas and the related micro-consciousnesses have been generated. According to this view, the binding results in the process that leads to join micro-consciousnesses together into a macro-consciousness. In Zeki’s terms, a macro-consciousness is a consciousness of a percept that consists of more than one attribute. In this sense, a percept can include components coming from different sensory modalities, so as to constitute a new distinct perceptual entity, for example, a buzzing white helicopter overhead.

Fig. 30.1 *Plato's Academy*, marble mosaic panel, Pompeii, Italy (Image cropped by authors)



Thus, according to Zeki, a micro-consciousness is the consciousness of one simple attribute (i.e., color, motion, orientation, etc.), while a macro-consciousness is the consciousness of a compound of simple attributes. In turn, a collection of macro-consciousnesses should lead to a globally unified field of consciousness. The picture is similar to that of a mosaic, in which an image is gradually created by means of an assemblage of minute colored pieces of glass, stone, or other materials, which, by themselves, do not constitute a coherent scene (Fig. 30.1). Metaphorically speaking, the single mosaic pieces could be compared to the micro-consciousnesses, while an object formed of more than one piece (such as the philosophers' clothes in the mosaic of our example) could be compared to a macro-consciousness; finally, the global scene pictured in the mosaic (the philosophers, the tree, and the pillars) could be compared to a unified conscious experience.

According to Zeki's hypothesis, the brain can engender multiple states of consciousness which constitute a hierarchy. At the top of this hierarchy, there would be a synthetic and unified consciousness, but this synthetic and unified conscious state would not be possible without other states of consciousness that lie at inferior levels of the information processing. As we have seen, Zeki identifies three hierarchical levels of consciousness: (1) the level of *micro-consciousness*, (2) the level of *macro-consciousness*, and (3) the level of the *unified consciousness*. Necessarily, each level depends on the existence of the previous one.

A logical consequence of Zeki's theory of multiple consciousnesses is that the greater the number of attributes to bind, the longer will be the time needed to produce a macro-consciousness and, eventually, a unified field of consciousness. In other words, binding between attributes (to form a macro-consciousness) takes more time than binding within attributes (to form a micro-consciousness). Although

neuroscientific findings are in line with this observation, more empirical evidence might be needed to support Zeki's view. In fact, it could be argued that neuroscientific evidence to date can equally support the theories that favor a global approach to consciousness. According to this approach, which is radically different to Zeki's view, consciousness is a single unified property generated by widespread networks in the brain. This widespread system is supposed to be formed of abundant thalamo-cortical reentrant connections, so as to create a global workspace within which information can be consciously elaborated (see Chaps. 16 and 19).

Despite the evidence suggesting that consciousness has to begin in the brain perceptual areas, it could be argued that these sites are necessary but not sufficient for elaborating some contents of phenomenal consciousness. In other words, high neuronal activity in V4 might be essential in order to consciously see color, not because V4 is capable to generate a micro-consciousness of color, but just because the color visual input must be elaborated unconsciously in V4 before entering the global neuronal workspace network of consciousness. Thus, differences of neuronal activity in a perceptual brain site might not be interpreted as a sign which distinguishes between unconscious and conscious processing, but rather as a sign which tells whether or not brain specialized areas are currently recruited into the global neuronal workspace network.

Moreover, Zeki's theory of multiple consciousnesses could lead to an unjustified proliferation of micro-consciousnesses. In fact, given that any brain processing area should in principle constitute a neural correlate of consciousness (NCC), the number of micro-consciousnesses could be equal to the number of brain processing areas. Still, certain brain regions, such as the cerebellum and some subcortical structures (e.g., the basal ganglia), do not appear to be necessarily involved in conscious processing and, thereby, do not appear to generate micro-consciousnesses per se. Therefore, Zeki's view needs to be further elaborated in order to explain why the activity of certain neurons correlates well with conscious experience, while that of others does not.

Finally, another problem for the theory of the multiple consciousnesses is that in certain experimental paradigms, it has been shown that visual regions are able to bind several attributes in the absence of consciousness, thus suggesting that the binding mechanism (which is supposed to produce macro-consciousnesses) can instead operate in the absence of consciousness. This empirical evidence challenges the claim that binding is a post-conscious mechanism.

Essential Bibliography

- Zeki S (2007) A theory of micro-consciousness. In: Velmans M, Schneider S (eds) *The Blackwell companion to consciousness*, Blackwell, Oxford, pp 580–588. This is a comprehensive summary of Zeki's theory of multiple consciousnesses. The author claims that consciousness is a composite and distributed property in time and space according to a rigid temporal hierarchy.

- Zeki S (2003) The disunity of consciousness. *Trends Cogn Sci* 7:214–218.
This article presents a clear introduction to the hierarchical model of consciousness. Neuroscientific evidence is put forward in order to support the composite nature of our conscious experience. The author suggests the hypothesis that the construction of a global perceptual scene, in which different attributes are consistently joined together (i.e., perceptual binding), might be a post-conscious phenomenon.

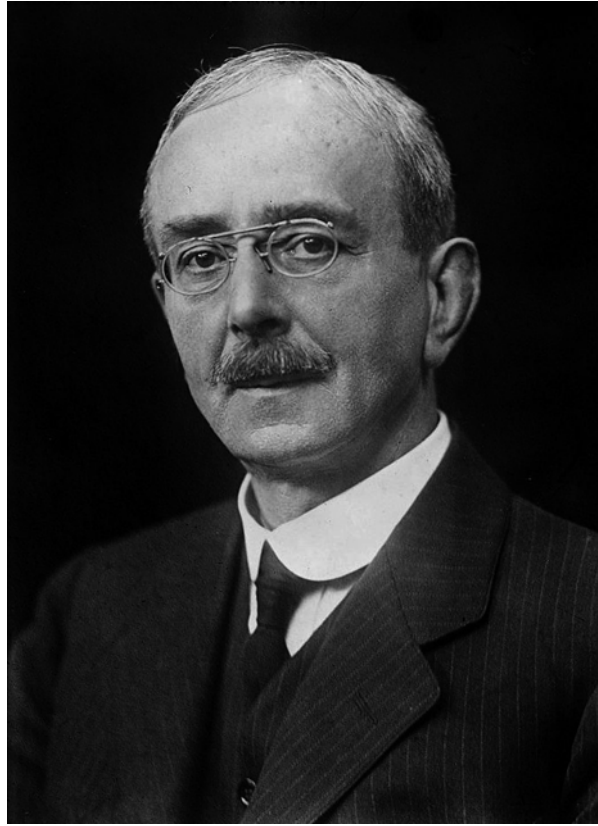
Epilogue: A Brief Tour of the Introductions to Consciousness Studies

That a man should simply and profoundly say that he cannot understand how consciousness has come into existence – is perfectly natural. But that a man should glue his eye to a microscope and stare and stare – and still not be able to see how it happens – is ridiculous, and it is particularly ridiculous when it is supposed to be serious. [...] If the natural scientists had been developed in Socrates' day as they are now, all the sophists would have been scientists. One would have hung a microscope outside his shop in order to attract custom, and then would have had a sign painted saying: "Learn and see through a giant microscope how a man thinks" (Soren Kierkegaard, diary entry of 1846)

Like a fortress under siege, consciousness has so far resisted any attempts to provide convincing solutions to the many questions it poses. The great neurophysiologist Sir Charles Sherrington (Fig. 1) neatly captured the problem of understanding the very nature of consciousness as a phenomenon which appears to elude scientific explanation within the domain of physical science: “The energy-concept [...] embraces and unifies much. [...] Immense as it is, and self-satisfying as it is, and self-contained as it is, it yet seems but an introduction to something else. For instance a star which we perceive. The energy-scheme deals with it, describes the passing of radiation thence into the eye, the little light-image of it formed at the bottom of the eye, the ensuing photo-chemical action in the retina, the trains of action-potentials travelling along the nerve to the brain, the further electrical disturbance in the brain, the action-potentials streaming thence to the muscles of eye-balls and of the pupil, the contraction of them sharpening the light-image and placing the best seeing part of the retina under it. The best ‘seeing’? That is where the energy-scheme forsakes us. It tells us nothing of any ‘seeing’. Everything but that. Of the physical happenings, yes. [...] But, as to our seeing the star it says nothing. [...] The energy-scheme deals with the star as one of the objects observable by us; as to the perceiving of it by the mind the scheme puts its finger to its lip and is silent. It may be said to bring us to the threshold of the act of perceiving, and there to bid us ‘goodbye’” (*Man On His Nature*, 1942, p. 304–305).

The view that the fundamental aspect of mental life is conscious experience which physical science cannot explain has an old tradition. To quote just a few examples, in the eighteenth century German philosopher Gottfried Leibniz (Fig. 2) proposed the famous analogy between the brain and the mill: “Suppose that there be

Fig. 1 Sir Charles Sherrington (1857–1952)



a machine, the structure of which produces thinking, feeling, and perceiving; imagine this machine enlarged but preserving the same proportions, so that you can enter it as if it were a mill. This being supposed, you might visit it inside; but what would you observe there? Nothing but parts which push and move each other, and never anything that could explain perception” (*Monadology*, Section XVII, 1714). Over a century later, British biologist Thomas Huxley (Fig. 3) wrote along the same lines “How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin, when Aladdin rubbed his lamp.”

More recently, German biologist Gunther Stent seemed to imply that the scientific exploration of consciousness should surrender, as the enterprise is to fail because the nature of consciousness is beyond the limits of human understanding. In a 1968 article published in *Science* which shares similarities with Colin McGinn’s new mysterianism (see Chap. 10), Stent wrote that “Searching for a “molecular” explanation of consciousness is a waste of time, since the physiological processes responsible for this wholly private experience will be seen to degenerate into seemingly

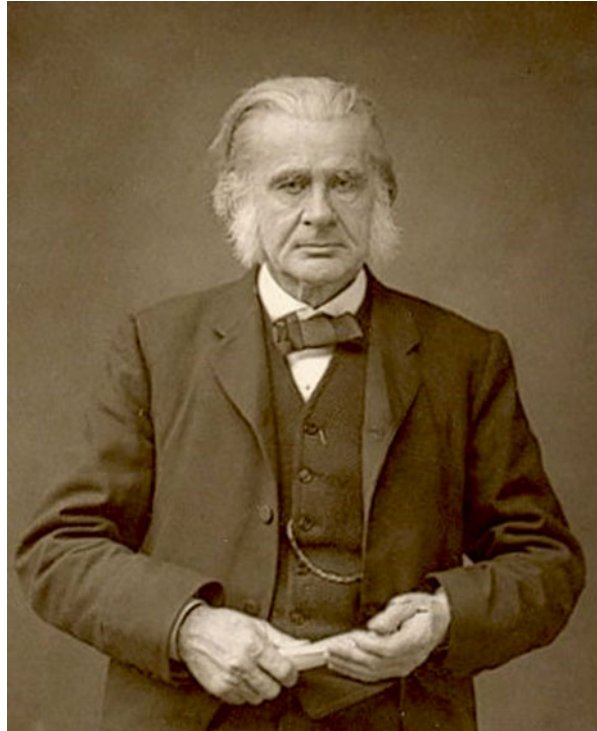
Fig. 2 Gottfried Leibniz (1646–1716)



quite ordinary, workaday reactions, no more and no less fascinating than those that occur in, say, the liver.” This pessimistic view about the possibility of reaching a scientific understanding of consciousness mirrors the strong ontological status, eliminating the problem altogether and relegating consciousness to a useless epiphenomenon of ontologically stronger physical entities (see Chaps. 2, 5, and 8). This latter view was elegantly epitomized by Thomas Huxley in his 1874 essay *On the hypothesis that animals are automata, and its history*: “The consciousness of brutes would appear to be related to the mechanism of their body simply as collateral product of its working, and to be completely without any power of modifying that working, as the steam-whistle which accompanies the work of a locomotive engine is without influence upon its machinery. Their volition, if they have any, is an emotion indicative of physical changes, not a cause of such changes [...] The soul stands to the body as the bell of a clock to the works, and consciousness answers to the sound which the bell gives out when it is struck [...] To the best of my judgment, the argumentation which applies to brutes holds good of men [...] We are conscious automata.”

The same feeling of frustration about the failed attempts to either decipher or deflate the problem of consciousness has not prevented the flourishing of a rich literature on the topic. The perceived importance – and intrinsic beauty – of the

Fig. 3 Thomas Huxley
(1825–1895)



conundrum of consciousness continues to generate both philosophical inquiry and scientific exploration across the mind-brain interface. In fact, the last three decades have seen an unprecedented increase in the number of publications by scholars of different disciplines. In turn, this has led to the development of the need for accessible introductory textbooks to such a vast and heterogeneous material, which we will briefly review in this final chapter (Fig. 4).

Introductory books on the newly established field of philosophy of mind first appeared over three decades ago, with Colin McGinn's *The Character of Mind: An Introduction to the Philosophy of Mind* (1982; second edition 1997), Paul Churchland's *Matter and Consciousness: Contemporary Introduction to the Philosophy of Mind* (1984; second edition 1988; third edition 2013), Owen Flanagan's *The Science of the Mind* (1984; second edition 1991), Peter Smith and Owen Jones' *The Philosophy of Mind: An Introduction* (1986), and William Bechtel's *Philosophy of Mind: An Overview for Cognitive Science* (1988). These books have soon become classics and, in many cases, have been updated in subsequent successful editions. The real "explosion" in the introductory literature to philosophy of mind began in the 1990s, with a series of excellent volumes which soon became popular as university course textbooks. These included George Graham's *Philosophy of Mind: An Introduction* (1993; second edition 1998), Dale Jacquette's *The Philosophy of Mind: The Metaphysics of Consciousness* (1994; second edition 2009), David Braddon-Mitchell and Frank Jackson's *Philosophy of Mind and Cognition* (1996; second edition 2006), Jaegwon Kim's *Philosophy of Mind* (1996; second edition 2006; third

Fig. 4 Detail from *The reader of novels* (1853) by Antoine Wiertz (1806–1865). Image cropped by authors



edition 2010), Sanford Goldberg and Andrew Pessin's *Gray Matters: An Introduction to the Philosophy of Mind* (1997), Colin McGinn's *Minds and Bodies: Philosophers and Their Ideas* (1997), Georges Rey's *Contemporary Philosophy of Mind: A Contentiously Classical Approach* (1997), Stephen Burwood et al.'s *Philosophy of Mind* (1998), and John Heil's *Philosophy of Mind* (1998; second edition 2004; third edition 2012). The new millennium has seen the consolidation of philosophy of mind as one of the most popular and rapidly expanding disciplines in academic departments of philosophy, mirrored by an unprecedented proliferation of introductory publications. In addition to updated editions of classic textbooks, newly published excellent books included Suzanne Cunningham's *What Is a Mind? An Integrative Introduction to the Philosophy of Mind* (2000), Samuel Guttenplan's *Mind's Landscape: An Introduction to the Philosophy of Mind* (2000), Jonathan Lowe's *An Introduction to the Philosophy of Mind* (2000), Andrew Brook and Robert Stainton's *Knowledge and Mind: A Philosophical Introduction* (2001), Tim Crane's *Elements of mind: An Introduction to the Philosophy of Mind* (2001), William Lyons's *Matters of the Mind* (2001), Keith Maslin's *An Introduction to the Philosophy of Mind* (2001; second edition 2007), Mel Thompson's *Understand Philosophy of Mind* (2003; second edition 2012), Neil Campbell's *A Brief Introduction to the Philosophy of Mind* (2005), Edward Feser's *Philosophy of Mind: A Short Introduction* (2005; second edition 2007), Ian Ravenscroft's *Philosophy of Mind: A Beginner's Guide* (2005), Barbara Montero's *On the Philosophy of Mind* (2009), David Cockburn's *An Introduction to the Philosophy of Mind: Souls, Science and Human Beings* (2011), William Jaworski's *Philosophy of Mind: A Comprehensive Introduction* (2011), Pete Mandik's *This Is Philosophy of Mind: An Introduction* (2013), and Andrew Bailey's (editor) *Philosophy of Mind: The Key Thinkers* (2014). Some of the most important academic publishers have also produced useful reference guides, such as Richard Gregory's (editor) *The Oxford Companion to the Mind* (1987; second edition 2004); Samuel Guttenplan's (editor) *A Companion to the Philosophy of Mind* (1996), Stephen Stich and Ted Warfield's (editors) *The Blackwell Guide to Philosophy of Mind* (2003), and James Garvey's (editor) *The Continuum Companion to Philosophy of Mind* (2011). Since the 1990s, introductory textbooks and compendia have been complemented by a number of comprehensive anthologies of classical and

contemporary readings in philosophy of mind, including William Lycan's (editor) *Mind and Cognition: An Anthology* (1990; second edition 1999; third edition 2008), David Rosenthal's (editor) *The Nature of Mind* (1991), Brian Beakley and Peter Ludlow's (editors) *The Philosophy of Mind: Classical Problems/Contemporary Issues* (1992; second edition 2006), Richard Warner and Tadeusz Szubka's (editors) *The Mind-Body Problem: A Guide to the Current Debate* (1994), William Lyons' (editor) *Modern Philosophy of Mind* (1995), Daniel Robinson's (editor) *The Mind* (1998), Anthony O'Hear's (editor) *Contemporary Issues in Philosophy of Mind* (1998), Peter Morton's (editor) *A Historical Introduction to the Philosophy of Mind: Readings with Commentary* (2000), David Chalmers' (editor) *Philosophy of Mind: Classical and Contemporary Readings* (2002), and John Heil's (editor) *Philosophy of Mind: A Guide and Anthology* (2004). Finally, Patrick Grim's (editor) *Mind and Consciousness: Five Questions* (2009) is a collection of interviews with leading philosophers of mind.

The mind-body problem is arguably the dominant theme within philosophy of mind, and – in Thomas Nagel's words – “Consciousness is what makes the mind-body problem really intractable” (*What is it like to be a bat?*, 1979). It is therefore not surprising that introductory textbooks specifically focusing on the philosophical literature on consciousness made their appearance since the turn of the new millennium, with William Seager's *Theories of consciousness: An introduction and assessment* (1999), David Papineau and Howard Selina's *Introducing Consciousness* (2000; second edition 2005), Max Velmans' *Understanding Consciousness* (2000; second edition 2009), Arne Dietrich's *Introduction to Consciousness* (2007), and Torin Alter and Robert Howell's *A Dialogue on Consciousness* (2009). Reference textbooks on the philosophical approach to the problem of consciousness are Max Velmans and Susan Schneider's (editors) *The Blackwell Companion to Consciousness* (2007), Philip Zelazo et al.'s (editors) *The Cambridge Handbook of Consciousness* (2007), and Tim Bayne et al.'s (editors) *The Oxford Companion to Consciousness* (2009). Ned Block et al.'s (editors) *The Nature of Consciousness: Philosophical Debates* (1997) and Torin Alter and Robert Howell's (editors) *Consciousness and the Mind-Body Problem: A Reader* (2012) are two useful anthologies of the most relevant philosophical articles on consciousness.

Over the last decades, philosophers have also turned their attention to the neurosciences, as shown by the publication of relevant books on the theoretical aspects of brain sciences. These include Maxwell Bennett and Peter Hacker's *Philosophical Foundations of Neuroscience* (2003), Maxwell Bennett et al.'s *Neuroscience and Philosophy: Brain, Mind, and Language* (2007), and Maxwell Bennett and Peter Hacker's *History of Cognitive Neuroscience* (2012), plus the anthologies by Ned Block (editor) *Readings in Philosophy of Psychology* (1980) and by William Bechtel et al. (editors) *Philosophy and the Neurosciences: A Reader* (2001).

The number of introductory books on the scientific approach to the problem of consciousness is considerably smaller compared to the philosophical literature; however, over the last decade, there have been a few relevant volumes, suggesting that consciousness studies have recently become a respectable research area within the neuroscientific tradition. Interestingly, the second edition of *Human Brain Function* (2004) with Richard Frackowiak as editor in chief contains a chapter

Fig. 5 Hippocrates of Kos
(450–380 BC)



entitled “The neural correlates of consciousness,” which opens with a remarkable sentence: “In the first edition of this book there was a final chapter on the future of imaging in which use of the word *consciousness* was strictly avoided until the very last sentence. Now that we have moved into a new millennium it has no longer been so easy to resist the *Zeitgeist*. That single sentence has become a whole chapter.” In a way, this paradigm shift toward neurobiological reductionism in consciousness studies had been anticipated by Hippocrates’ writings around 400 years before the Christian era (Fig. 5): “Men ought to know that from the brain, and from the brain only arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant” (*On the Sacred Disease*). In more recent times, the same concept was poetically reinstated by George Bernard Shaw at the beginning of the last century in the form of a dialogue between Don Juan and the Devil: “*The Devil*: You conclude, then, that Life was driving at clumsiness and ugliness? *Don Juan*: No, perverse devil that you are, a thousand times no. Life was driving at brains – at its darling object: an organ by which it can attain not only self-consciousness but self-understanding” (Shaw, *Man and Superman: Don Juan in Hell, Act III*, 1903). The belief that the study of brain function could shed light on the nature of consciousness was also echoed in



Fig. 6 Michelangelo Buonarroti (1475–1564), *The Creation of Adam* (1511–1512), fresco, Sistine Chapel, Vatican. Different authors have recently suggested that the shape of God portrayed in the act of giving reason to Adam corresponds to the sagittal section of a human brain (Meshberger 1990; Paluzzi et al. 2007; Suk and Tamargo 2010)

Somerset Maugham’s words before it became a self-fulfilling prophecy (Fig. 6): “The highest activities of consciousness have their origins in the physical occurrences of the brain just as the loveliest melodies are not too sublime to be expressed by notes” (Maugham, *A Writer’s Notebook*, 1949).

The novel stream of neuroscientific publications on consciousness includes Adam Zeman’s *Consciousness: A User’s Guide* (2004), Bernard Baars and Nicole Gage’s *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience* (2007; second edition 2010), Steven Laureys and Giulio Tononi’s (editors) *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology* (2008), Andrea Cavanna et al.’s (editors) *Neuroimaging of Consciousness* (2013), and Bernard Baars et al.’s *Essential Sources in the Scientific Study of Consciousness* (2003). A couple of excellent articles published in two of the most important scientific journals in the medical disciplines of neurology and psychiatry are Adam Zeman’s “Consciousness” (*Brain* 2001;124:1263–1289) and Kenneth Kendler’s “A psychiatric dialogue on the mind-body problem” (*American Journal of Psychiatry* 2001;158:989–1000).

Our brief tour of the introductory literature to consciousness studies closes with an inspiring handful of books, which have made the first innovative attempts to bridge the “two cultures,” by devoting roughly equal space and attention to both philosophical and neuroscientific theories. These books, which share the multidisciplinary spirit of the present volume, include Thomas Metzinger’s

(editor) *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (2000), Stanislas Dehaene's (editor) *The Cognitive Neuroscience of Consciousness* (2001), Rita Carter's *Exploring Consciousness* (2002; second edition 2004), Susan Blackmore's *Consciousness: An Introduction* (2003; second edition 2010) and *Consciousness: A Very Short Introduction* (2005a), David Rose's *Consciousness: Philosophical, Psychological, and Neural Theories* (2006), and Antti Revonsuo's *Consciousness: The Science of Subjectivity* (2009). Susan Blackmore also published a lively and insightful collection of interviews with both philosophers and scientists of consciousness, titled *Conversations on Consciousness* (2005a, b; second edition 2007).

The digital era offers useful resources to the interested readers who wish to keep up to date with the rapidly expanding philosophical and neuroscientific literature on consciousness. Of particular relevance are a few interdisciplinary peer-reviewed academic journals dedicated entirely to the field of consciousness studies: *Consciousness and Cognition* (founded in 1992: <http://www.journals.elsevier.com/consciousness-and-cognition/>), *Journal of Consciousness Studies* (founded in 1994: <http://ingentaconnect.com/journals/browse/imp/jcs>), and *Frontiers in Consciousness Research* (founded in 2010: http://www.frontiersin.org/consciousness_research). *Psyche* (http://www.theassc.org/journal_psyche) is a free electronic journal dedicated to supporting the interdisciplinary exploration of the nature of consciousness and its relation to the brain that was active between 1994 and 2010. The journal *Consciousness and Emotion* (founded in 2000: <https://benjamins.com/#catalog/journals/ce/main>) in 2003 became a book series (<https://benjamins.com/#catalog/books/ceb/main>), alongside *Advances in Consciousness Research* (founded in 1995: <https://benjamins.com/#catalog/books/aicr/main>).

A few precious repositories of online resources need mentioning. The *Stanford Encyclopedia of Philosophy* has a wonderful website (<http://plato.stanford.edu/>), which organizes scholars from around the world in philosophy and related disciplines to create and maintain an up-to-date reference work. The philosophy of mind section of the website contains several entries relevant to consciousness studies which are highly accurate and authoritative. David Chalmers' website (<http://consc.net/chalmers/>) is arguably the most comprehensive collection on topics related to consciousness and/or philosophy, such as the bibliography MindPapers, directories of online papers, and some philosophical diversions, including a photo gallery. US-based Italian polymath Piero Scaruffi has single-handedly compiled an annotated bibliography of mind-related topics in his superb website (<http://www.scaruffi.com/mind.html>), a real gold mine for both experts and beginners.

The Center for Consciousness Studies based at the University of Arizona (<http://www.consciousness.arizona.edu/>) promotes a series of successful meetings called Towards a Science of Consciousness (TSC: <https://sbs.arizona.edu/project/consciousness/>). The TSC conferences are the preeminent world gatherings on all approaches to the profound and fundamental question of how the brain produces conscious experience, a question which addresses who we are, the nature of reality, and our place in the universe. These interdisciplinary conferences emphasize broad and rigorous approaches to all aspects of the study and understanding of conscious

awareness. Topical areas include neuroscience, philosophy, psychology, biology, quantum physics, meditation and altered states, machine consciousness, culture, and experiential phenomenology. Held annually since 1994, these conferences alternate yearly between Tucson, Arizona, and various locations around the world.

The Association for the Scientific Study of Consciousness (ASSC: <http://www.theassc.org/>) is an academic society that promotes rigorous research directed toward understanding the nature, function, and underlying mechanisms of consciousness. The ASSC includes members working in the fields of cognitive science, medicine, neuroscience, and philosophy, along with other relevant disciplines in the sciences and humanities, and coordinates a series of successful annual conferences on the scientific study of consciousness which started in 1997.

We hope that these colorful, albeit partial, snapshots of introductory books and online resources on consciousness studies will serve the reader as an Ariadne's thread to navigate the infinite passages of the consciousness labyrinth. In a way, they highlight the value of interdisciplinarity as the most fruitful way forward in this complex field. Inspired by the genuine fascination of the subject, we wish the readers good luck in their journeys through the delightful mysteries of human consciousness.

Bibliography

- Alter T, Howell RJ (2009) A dialogue on consciousness. Oxford University Press, Oxford
- Alter T, Howell RJ (eds) (2012) Consciousness and the mind-body problem: a reader. Oxford University Press, Oxford
- Baars BJ (1988) A cognitive theory of consciousness. Cambridge University Press, Cambridge, MA
- Baars BJ (1997a) In the theater of consciousness: the workspace of the mind. Oxford University Press, Oxford
- Baars BJ (1997b) In the theater of consciousness: global workspace theory, a rigorous scientific theory of consciousness. *J Conscious Stud* 4:292–309
- Baars BJ, Gage NM (2007) Cognition, brain, and consciousness: introduction to cognitive neuroscience. Academic Press, London, 2nd edn 2010
- Baars BJ, Banks WP, Newman JB (2003) Essential sources in the scientific study of consciousness. MIT Press, Cambridge, MA
- Baddeley A (2007) Working memory, thought, and action. Oxford University Press, Oxford
- Bailey A (ed) (2014) Philosophy of mind: the key thinkers. Bloomsbury, London
- Bayne T, Cleeremans A, Wilken P (eds) (2009) The Oxford companion to consciousness. Oxford University Press, Oxford
- Beakley B, Ludlow P (eds) (1992) The philosophy of mind: classical problems/Contemporary issues. MIT Press, Cambridge, MA, 2nd edn 2006
- Bechtel W (1988) Philosophy of mind: an overview for cognitive science. Lawrence Erlbaum, Hillsdale
- Bechtel W, Mandik P, Mundale J, Stufflebeam RS (eds) (2001) Philosophy and the neurosciences: a reader. Wiley-Blackwell, Oxford
- Bekoff M, Sherman PW (2004) Reflections on animal selves. *Trends Ecol Evol* 19:176–180
- Bennett MR, Hacker PMS (2003) Philosophical foundations of neuroscience. Wiley-Blackwell, Oxford
- Bennett MR, Hacker PMS (2012) History of cognitive neuroscience. Wiley-Blackwell, Oxford
- Bennett M, Dennett D, Hacker P, Searle J (2007) Neuroscience and philosophy: brain, mind, and language. Columbia University Press, New York
- Blackmore S (2003) Consciousness: an introduction. Hodder and Stoughton, London, 2nd edn 2010
- Blackmore S (2005a) Consciousness: a very short introduction. Oxford University Press, Oxford
- Blackmore S (2005b) Conversations on consciousness. Oxford University Press, Oxford, 2nd edn 2007
- Block N (ed) (1980) Readings in philosophy of psychology. Harvard University Press, Cambridge, MA
- Block N (1996) On a confusion about a function of consciousness. *Behav Brain Sci* 18:227–247
- Block N, Flanagan O, Guzeldere G (eds) (1997) The nature of consciousness: philosophical debates. MIT Press, Cambridge, MA

- Braddon-Mitchell D, Jackson F (1996) *Philosophy of mind and cognition*. Wiley-Blackwell, Oxford, 2nd edn 2006
- Brook A, Stainton RJ (2001) *Knowledge and mind: a philosophical introduction*. MIT Press, Cambridge, MA
- Burwood S, Gilbert P, Lennon K (1998) *Philosophy of mind*. McGill Queens University Press, Montreal
- Campbell N (2005) *A brief introduction to the philosophy of mind*. Broadview, Peterborough
- Carter R (2002) *Exploring consciousness*. University of California Press, Berkeley, 2nd edn 2004
- Cavanna AE, Nani A, Blumenfeld H, Laureys S (eds) (2013) *Neuroimaging of consciousness*. Springer, Berlin
- Chalmers D (1995) Facing up to the problem of consciousness. *J Consc Stud* 2:200–219
- Chalmers D (1996) *The conscious mind: in search of a fundamental theory*. Oxford University Press, New York
- Chalmers D (ed) (2002) *Philosophy of mind: classical and contemporary readings*. Oxford University Press, Oxford
- Chalmers D (2010) *The character of consciousness*. Oxford University Press, New York
- Churchland PM (1984) *Matter and consciousness: contemporary introduction to the philosophy of mind*. MIT Press, Cambridge, MA, 2nd edn 1988; 3rd edn 2013
- Churchland PS (2002) *Brain-wise: studies in neurophilosophy*. MIT Press, Cambridge, MA
- Churchland PS (2011) *Braintrust: what neuroscience tells us about morality*. Princeton University Press, Princeton
- Churchland PM (2012) *Plato's camera: how the physical brain captures a landscape of abstract universals*. MIT Press, Cambridge, MA
- Churchland PM (2013) *Matter and consciousness*, 3rd edn. MIT Press, Cambridge, MA
- Cockburn D (2011) *An introduction to the philosophy of mind: souls, science and human beings*. Pelgrave Macmillan, Basingstoke
- Crane T (2001) *Elements of mind: an introduction to the philosophy of mind*. Oxford University Press, New York
- Crane T (2003) *The mechanical mind: a philosophical introduction to minds, machines and mental representation*. Routledge, London
- Crane T (2014) *Aspect of psychologism*. Harvard University Press, Cambridge, MA
- Crick F (1994) *The astonishing hypothesis: the scientific search for the soul*. Charles Scribner's Sons, New York
- Cunningham S (2000) *What is a mind? An integrative introduction to the philosophy of mind*. Hackett, Indianapolis
- Damasio A (1994) *Descartes' error: emotion, reason, and the human brain*. Putnam, New York
- Damasio A (1999) *The feeling of what happens: body and emotion in the making of consciousness*. Harcourt, Orlando
- Damasio A (2010) *Self comes to mind: constructing the conscious brain*. Pantheon Books, New York
- Davidson D (1985) Reply to Quine on Events. In: LePore E, McLaughlin B (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Blackwell, New York
- Davidson D (1993) *Thinking Causes*. In: Heil J, Mele A (eds.), *Mental Causation*, Clarendon Press, Oxford
- Davidson D (1995) *Laws and Cause*. *Dialectica* 49(2–4):263–279
- Davidson D (2001a) *Essays on actions and events*, 2nd edn. Clarendon, Oxford
- Davidson D (2001b) *Subjective, intersubjective, objective*. Clarendon, Oxford
- Davidson D (2004) *Problems of rationality*. Clarendon, Oxford
- Dehaene S (ed) (2001) *The cognitive neuroscience of consciousness*. Elsevier, Amsterdam
- Dehaene S (2014) *Consciousness and the brain: deciphering how the brain codes our thoughts*. Viking, New York
- Dehaene S, Naccache L (2001) *Towards a cognitive neuroscience of consciousness: basic evidence and a workspace network*. *Cognition* 79:1–37
- Dennett DC (1987) *The intentional stance*. MIT Press, Cambridge, MA

- Dennett DC (1991) *Consciousness explained*. Little, Brown, Boston, MA
- Dennett DC (2005) *Sweet dreams: philosophical obstacles to a science of consciousness*. MIT Press, Cambridge, MA
- Descartes R (1637) Discours de la méthode pour bien conduire sa raison & chercher la vérité dans les sciences. Plus la dioptrique. Les météores. Et la géométrie. Qui sont des essais de cette Méthode. In: *Œuvres de Descartes*, vol 11. Charles Adam and Paul Tannery, Vrin/CNRS, Paris; English translation: *Discourse on the method*. In: Cottingham J, Stoothoff R, Murdoch D Kenny A (1984–1991) *Philosophical writings of Descartes*, vol 3. Cambridge University Press, Cambridge
- Descartes R (1641) *Meditationes de prima philosophia*. In: *Œuvres de Descartes*, vol 11. edn. Charles Adam and Paul Tannery. Vrin/CNRS, Paris. English translation: *Meditations on first philosophy*. In: Cottingham J, Stoothoff R, Murdoch D, Kenny A (1984–1991) *Philosophical writings of Descartes*, 3 vol. Cambridge University Press, Cambridge
- Descartes R (1649) *Les passions de l'âme*. In: *Œuvres de Descartes*, vol 11. edn. Charles Adam and Paul Tannery. Vrin/CNRS, Paris. English translation: *Passions of the soul*. In Cottingham J, Stoothoff R, Murdoch D, and Kenny A (1984–1991) *Philosophical writings of Descartes*, vol 3. Cambridge University Press, Cambridge
- Dietrich A (2007) *Introduction to consciousness*. Macmillan, London
- Donald M (1991) *Origin of the modern mind: three stages in the evolution of culture and cognition*. Harvard University Press, Cambridge, MA
- Donald M (2001) *A mind so rare: the evolution of human consciousness*. W W Norton & Company Incorporated, New York
- Eccles JC (1989) *Evolution of the brain: creation of the self*. Routledge, London
- Edelman GM (1989) *The remembered present: a biological theory of consciousness*. Basic Books, New York
- Edelman GM (1992) *Bright air, brilliant fire: on the matter of the mind*. Basic Books, New York
- Edelman GM (2007) *Second nature: brain science and human knowledge*. Yale University Press, New Haven
- Feser E (2005) *Philosophy of mind: a short introduction*. Oneworld, Oxford, 2nd edn 2007
- Flanagan O (1984) *The science of the mind*. MIT Press, Cambridge, MA, 2nd edn 1991
- Fodor J (1983) *The modularity of mind*. MIT Press, Cambridge, MA
- Fodor J (1994) *The elm and the expert: mentalese and its semantics*. MIT Press, Cambridge, MA
- Fodor J (2000) *The mind doesn't work that way: the scope and limits of computational psychology*. MIT Press, Cambridge, MA
- Fodor J (2008) *LOT 2: the language of thought revisited*. Oxford University Press, Oxford
- Foster J (1991) *The immaterial self: a defence of the Cartesian dualist*. *Conception of the Mind*. Routledge, London
- Garber D (2001) *Descartes embodied*. Cambridge University Press, Cambridge
- Garvey J (ed) (2011) *The Continuum companion to philosophy of mind*. Continuum, London
- Goldberg S, Pessin A (1997) *Gray matters: an introduction to the philosophy of mind*. Sharpe, New York
- Graham G (1993) *Philosophy of mind: an introduction*. Wiley-Blackwell, Oxford, 2nd edn 1998
- Gregory RL (ed) (1987) *The Oxford companion to the mind*. Oxford University Press, Oxford, 2nd edn 2004
- Grim P (ed) (2009) *Mind and consciousness: five questions*. Automatic Press, London
- Guttenplan S (ed) (1996) *A companion to the philosophy of mind*. Wiley-Blackwell, Oxford
- Guttenplan S (2000) *Mind's landscape: an introduction to the philosophy of mind*. Wiley-Blackwell, Oxford
- Hameroff S, Penrose R (2014) *Consciousness in the universe: a review of the 'Orch OR' theory*. *Phys Life Rev* 11:39–78
- Hart WD (1988) *The engines of the soul*. Cambridge University Press, Cambridge
- Hasker W (1999) *The emergent self*. Cornell University Press, Ithaca
- Heil J (1998) *Philosophy of mind: a contemporary introduction*. Routledge, London, 2nd edn 2004; 3rd edn 2012

- Heil J (2004) *Philosophy of mind: a guide and anthology*. Oxford University Press, Oxford
- Humphrey N (1992) *A history of the mind: evolution and the birth of consciousness*. Chatto & Windus, London
- Humphrey N (2002) *The mind made flesh: essays from the frontiers of evolution and psychology*. Oxford University Press, Oxford
- Humphrey N (2006) *Seeing red: a study in consciousness*. Harvard University Press, Cambridge, MA
- Humphrey N (2008) *Getting the measure of consciousness*. *Prog Theor Phys Suppl* 173:264–269
- Humphrey N (2011) *Soul dust: the magic of consciousness*. Princeton University Press, Princeton
- Huxley TH (1866) *Lessons in elementary physiology*. Macmillan, London, p 193
- Huxley TH (1874) *On the hypothesis that animals are automata, and its history*. *Fortn Rev* 16:555–580
- Huxley TH (1884) *Animal automatism, and other essays*. The Humboldt Publishing Company, New York
- Jacquette D (1994) *The philosophy of mind: the metaphysics of consciousness*. Bloomsbury, London, 2nd edn 2009
- Jaworski W (2011) *Philosophy of mind: a comprehensive introduction*. Wiley-Blackwell, Oxford
- Jaynes J (1976) *The Origin of consciousness in the breakdown of the bicameral mind*. Houghton Mifflin, Boston, MA
- Jaynes J (1992) *The Julian Jaynes collection*. Basic Books, New York
- Kendler KS (2001) *A psychiatric dialogue on the mind-body problem*. *Am J Psychiatry* 158: 989–1000
- Kim J (1993) *Supervenience and mind*. Cambridge University Press, Cambridge
- Kim J (1996) *Philosophy of mind*. Westview, Boulder, 2nd edn 2006; 3rd edn 2010
- Kim J (1998) *Mind in a physical world*. MIT Press, Cambridge, MA
- Kim J (2005) *Physicalism or something near enough*. Princeton University Press, Princeton
- Koch C (2002) *The quest for consciousness*. Roberts & Company Publishers, Englewood
- Koch C (2012) *Consciousness: confessions of a romantic reductionist*. MIT Press, Cambridge, MA
- Kripke S (1980) *Naming and necessity*. Harvard University Press, Cambridge, MA
- Laureys S, Tononi G (eds) (2008) *The neurology of consciousness: cognitive neuroscience and neuropathology*. Academic, London
- Libet B (1999) *Do we have free will?* *J Consc Stud* 6:47–57
- Libet B (2004) *Mind time: the temporal factor in consciousness*. Harvard University Press, Cambridge, MA
- Lowe EJ (2000) *An introduction to the philosophy of mind*. Cambridge University Press, Cambridge
- Lowe EJ (2006) *The four-category ontology: a metaphysical foundation for natural science*. Oxford University Press, Oxford
- Lycan W (1987) *Consciousness*. MIT Press, Cambridge, MA
- Lycan W (ed) (1990) *Mind and cognition: an anthology*. Wiley-Blackwell, Oxford, 2nd edn 1999; 3rd edn 2008
- Lycan W (1996) *Consciousness and experience*. MIT Press, Cambridge, MA
- Lyons W (ed) (1995) *Modern philosophy of mind*. Orion, London
- Lyons W (2001) *Matters of the mind*. Routledge, London
- Mandik P (2013) *This is philosophy of mind: an introduction*. Wiley-Blackwell, Oxford
- Maslin KT (2001) *An introduction to the philosophy of mind*. Wiley-Blackwell, Oxford, 2nd edn 2007
- McGinn C (1982) *The character of mind: an introduction to the philosophy of mind*. Oxford University Press, Oxford, 2nd edn 1997
- McGinn C (1991) *The problems of consciousness*. Blackwell, Oxford
- McGinn C (1997) *Minds and bodies: philosophers and their ideas*. Oxford University Press, Oxford
- McGinn C (1999) *The mysterious flame: conscious minds in a material world*. Basic Book, New York

- McGinn C (2004) *Consciousness and its objects*. Oxford University Press, Oxford
- Meshberger FL (1990) An interpretation of Michelangelo's Creation of Adam based on neuro-anatomy. *JAMA* 265:1837–1841
- Metzinger T (ed) (2000) *Neural correlates of consciousness: empirical and conceptual questions*. MIT Press, Cambridge, MA
- Montero B (2009) *On the philosophy of mind*. Wadsworth, Belmont
- Morton P (ed) (2000) *A historical introduction to the philosophy of mind: readings with commentary*. Broadview, Peterborough
- Maugham S (1949) *A writer's notebook*. William Heinemann, London
- Nagel T (1979) What is it like to be a bat? *Philos Rev* LXXXIII:435–50; reprinted in: Nagel T (1979) *Mortal questions*. Cambridge University Press, Cambridge
- Nagel T (2000) The psychophysical nexus. In: Boghossian P, Peacocke C (eds) *New essays on the a priori*. Clarendon Press, Oxford; reprinted in Nagel T (2002) *Concealment and exposure & other essays*, Oxford University Press, Oxford
- Nagel T (2012) *Mind and cosmos*. Oxford University Press, Oxford
- Noë A (2004) *Action in perception*. MIT Press, Cambridge, MA
- Noë A (2009) *Out of our heads: why you are not your brain, and other lessons from the biology of consciousness*. Hill and Wang, New York
- Noë A (2012) *Varieties of presence*. Harvard University Press, Cambridge, MA
- O'Hear A (ed) (1998) *Contemporary issues in philosophy of mind*. Cambridge University Press, Cambridge
- O'Regan JK (2011) Why red doesn't sound like a bell: understanding the feel of consciousness. Oxford University Press, Oxford
- O'Regan JK (2012) How to build a robot that is conscious and feels. *Minds Mach* 22:117–136
- Paluzzi A, Belli A, Bain P, Viva L (2007) Brain 'imaging' in the Renaissance. *J R Soc Med* 100:540–543
- Papineau D, Selina H (2000) *Introducing consciousness*. Icon Books, Cambridge, 2nd edn 2005
- Penrose R (1989) *The emperor's new mind: concerning computers, minds, and the laws of physics*. Oxford University Press, Oxford
- Penrose R (1994) *Shadows of the mind: an approach to the missing science of consciousness*. Oxford University Press, Oxford
- Popper KR (1994) *Knowledge and the mind-body problem: in defence of interaction*. Routledge, London
- Popper KR, Eccles JC (1977) *The self and its brain: an argument for interactionism*. Springer, Berlin; reprinted in 2012, Springer London, Limited
- Portas C, Maquet P, Rees G, Blakemore S, Frith C (2004) The neural correlates of consciousness. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, Ashburner J, Penny W (eds) *Human brain function*. Academic, San Diego, pp 269–302
- Putnam H (1975) *Mind, language, and reality*. Philosophical papers, vol 2. Cambridge University Press, Cambridge
- Putnam H (1988) *Representation and reality*. The MIT Press, Cambridge, MA
- Putnam H (2001) *The threefold cord: mind, body, and world*. Columbia University Press, New York
- Ravenscroft I (2005) *Philosophy of mind: a beginner's guide*. Oxford University Press, Oxford
- Revonsuo A (2009) *Consciousness: the science of subjectivity*. Psychology Press, New York
- Rey G (1997) *Contemporary philosophy of mind: a contentiously classical approach*. Wiley-Blackwell, Oxford
- Robinson D (ed) (1998) *The mind*. Oxford University Press, Oxford
- Rose D (2006) *Consciousness: philosophical, psychological, and neural theories*. Oxford University Press, Oxford
- Rosenthal D (ed) (1991) *The nature of mind*. Oxford University Press, Oxford
- Rosenthal D (2005) *Consciousness and mind*. Clarendon, Oxford
- Rosenthal D (2008) Consciousness and its function. *Neuropsychologia* 46:829–840
- Seager W (1999) *Theories of consciousness: an introduction and assessment*. Routledge, London
- Searle JR (1992) *The rediscovery of the mind*. MIT Press, Cambridge, MA

- Searle JR (2002) *Consciousness and language*. Cambridge University Press, Cambridge
- Searle JR (2004) *Mind: a brief introduction*. Oxford University Press, Oxford
- Shaw GB (1903) *Man and superman*. Cambridge University Press, Cambridge
- Sherrington C (1942) *Man on his nature*. Cambridge University Press, Cambridge
- Smith P, Jones OR (1986) *The philosophy of mind: an introduction*. Cambridge University Press, Cambridge
- Stent GS (1968) That was the molecular biology that was. *Science* 160:390–395
- Stich SP, Warfield TA (eds) (2003) *The Blackwell guide to philosophy of mind*. Wiley-Blackwell, Oxford
- Suk I, Tamargo RJ (2010) Concealed neuroanatomy in Michelangelo's Separation of Light From Darkness in the Sistine Chapel. *Neurosurgery* 66:851–861
- Swinburne R (1997) *The evolution of the soul*. Oxford University Press, Oxford
- Thompson M (2003) *Understand philosophy of mind*. Hodder and Stoughton, London, 2nd edn 2012
- Tononi G (2004) An information integration theory of consciousness. *BMC Neurosci* 5:42
- Tononi G (2011) Consciousness as integrated information: a provisional manifesto. *Biol Bull* 215:216–242
- Tononi G (2012) *Phi: a voyage from the brain to the soul*. Pantheon Books, New York
- Velmans M (2000) *Understanding consciousness*. Routledge, London, 2nd edn 2009
- Velmans M (2008) Reflexive monism. *J Consc Stud* 15:5–50
- Velmans M (2009) *Understanding consciousness*. Routledge, London
- Velmans M, Schneider S (eds) (2007) *The Blackwell companion to consciousness*. Blackwell, Oxford
- Warner R, Szubka T (eds) (1994) *The mind-body problem: a guide to the current debate*. Wiley-Blackwell, Oxford
- Wittgenstein L (2001) *Philosophical investigations*. Blackwell, Oxford
- Zeki S (2003) The disunity of consciousness. *Trends Cogn Sci* 7:214–218
- Zeki S (2007) A theory of micro-consciousness. In: Velmans M, Schneider S (eds) *The Blackwell companion to consciousness*. Blackwell, Oxford, pp 580–588
- Zelazo PD, Moscovitch M, Thompson E (eds) (2007) *The Cambridge handbook of consciousness*. Cambridge University Press, Cambridge
- Zeman A (2001) Consciousness. *Brain* 124:1263–1289
- Zeman A (2004) *Consciousness: a user's guide*. Yale University Press, New Haven

Index

A

Access consciousness, 96
Achilles slays Hector, 142
Achromatopsia, 176
Acquaintance, 31, 56–58
Akinetopsia, 176
A Mind So Rare, 117
Anomalous monism. *See* Davidson's
 anomalous monism
Argument of automaton, 5
Ariel on a Bat's Back, 62
Artificial consciousness, 28, 88, 115
The Art of Painting, 68, 69
Ascending reticular activating system
 (ARAS), 101
Aspectual shape, 17

B

Baars's workspace theory
 Cartesian theater, 94
 conscious access hypothesis, 95, 96
 global workspace theater, 93
 scientific inquiry, consciousness, 93
 The Teatro Regio in Turin, 94
Belvedere, 22, 23
The Bicameral mind. *See* Jaynes'
 theoretical analysis
Binding problem
 coherent single visual picture, 101
 micro- and macro-consciousness,
 176, 177
 reentry process, 129
 subject's awareness, events, 51–52
Biological naturalism. *See also* Searle's
 perspective
 building block theory, 87
 description, 87
 unified field theory, 87

The Blank Signature, 151
Brain web, 112–114
Building block theory, 87

C

Cartesian coordinate system, 29
Cartesian materialism, 26
Cartesian theater, 26, 94
Causal and ontological reducibility, 86
Causal explanations, 21
Causal overdetermination, 45, 126
Chalmers philosophy. *See also* Property
 dualism
 argument of automaton, 5
 brain activity, 5–6
 easy and hard problem, 3
 experiences, 4
 metaphysics and epistemology, 3
 naturalistic dualism/property dualism, 4
 principles, 6–7
Change blindness phenomenon, 151
“Chinese room” argumentation, 88
CMF. *See* Conscious mental field (CMF)
The Concept of Mind, 25
Conscious access hypothesis, 95, 96
Conscious mental field (CMF)
 experiences, 148
 nonphysical and nonreducible emergent
 property, 149
 phenomenological category, 148
The Conscious Mind, 3
Consciousness
 binding problem, 4, 52
 Descartes' hypothesis (*see* Descartes'
 hypothesis)
 easy problem, 3
 Fodor's hypothesis (*see* Fodor's
 hypothesis)

- Consciousness (*cont.*)
 functional approach, 28
 hard problem, 4
 intentionality, 15–18
 Kim's philosophy (*see* Kim's philosophy)
 neurophilosophy and eliminative
 materialism, 9–13
 property dualism, 6–7
 Conscious veto awareness, 147
 Control consciousness, 50
 Core consciousness, 109
 Crane's hypothesis
 intentionality, 15–18
 psychologism (*see* Psychologism)
 Creature consciousness, 79
 Crick and Koch contributions
 binding problem, 101, 102
 brain properties, 99–100
 conscious experience and neuronal
 processes, 100
The Harp of Erin, 101, 102
 midbrain/hindbrain structures, 101
 penumbra, neuronal activity, 103
 snapshot and threshold, 102
 zombie modes, 102
- D**
- Damasio theory
 autobiographical self, 107–109
 brain activity and cognitive function, 105
 core and extended consciousness, 109
 dynamic maps, 106
 emotions and feelings, 106
 ingredients, consciousness, 106
 perspective, ownership, agency and
 primordial feelings, 107
 protosef and core self, 107
 somatic marker hypothesis, 105
 Davidson's anomalous monism
 anomalism principle, 20
Belvedere, 22, 23
 causal explanations, 21
 cause-law principle, 20
 definition, 19
 "event", 21
 intentionality, 22
 interaction principle, 20
 ontology and epistemology, 19–20
 token and type identity theory, 20–21
 Dehaene's theoretical approach
 amplification, sensory brain activity, 112
 artificial consciousness, 115
 brain web, 112–114
 characteristic EEG pattern, 112
 evolutionary perspective, 114
 global ignition, 112
 neuronal global workspace, 112
 neuronal models, human cognition, 111
 phenomenal consciousness, 114
 Dendrons, 125
 Dennett's philosophical contributions
 heterophenomenology, 26
 intentional stance, 26
 multiple drafts model, 27
 qualia (*see* Qualia)
 Descartes' hypothesis
 Cartesian/classical dualism, 29–30, 35
Discourse on the Method, 30, 31
 dreams, 29
Meditations on First Philosophy, 31–33
Passions of the Soul, 31, 33
 pineal gland, location of, 34
 portrait, 29, 30
Discourse on the Method, 31
 Donald's hypothesis
 cognitive skills and reprogramming, 120
 hard problem, 121
 human mind development, 117
 memory representation, 118
 mimetic transition, 118
 mythic transition, 118, 119
 primary motor cortex, 119
 secondary and tertiary cortex, 120
 theoretic transition, 118
 Double-aspect theory of information, 7
- E**
- Easy problem, 3
 Edelman's theoretical framework
 cellular processes, 127
 general, informational and
 subjective, 129
 natural property, living human
 brain, 127
 neuronal group selection, 128
 postnatal synaptic modifications, 128
 processes, regulation, 128
 reentry process, 128, 129
 regulation mechanisms, genetic and
 epigenetic, 128
 thalamocortical system (*see*
 Thalamocortical system)
 Edelman's theory of neuronal group
 selection, 128, 131
 Eliminative materialism
 definition, 9
 folk psychology (*see* Folk psychology)
 and neurophilosophy, 9, 12–13
 opponents of, 10
 reflexive monism, 169

- Embodied cognition, 77
 Engram (recorded information on biological supports), 118
 Epiphenomenalism, 130
 Event consciousness, 50
 Exogram (recorded information on external supports), 118
 Explanatory gap, 4, 52
 Extended consciousness, 109
 Exteroceptive maps, 106
- F**
 First-person epistemology, 87
 First-person ontology, 86
Flora, portrait of a lady, 39, 40
 Fodor's hypothesis
 cognition, 39
 folk psychology, 37
 inferential role semantics, 39
 informational atomism, 39
 language of thought hypothesis, 38
 mental representational theory, 41
 propositional attitudes, 37, 38
 Folk psychology
 classification and causal generalizations, 10
 description, 9, 37
 phlogiston, 9–10
 propositional attitudes, 38
 Functional isomorphism
 cognitive science, 74
 description, 74
 embodied cognition, 77
 hydraulic pipelines metaphor, 74
 objections against, 76, 77
 pain and mental events, 74–75
 philosophical zombie argument, 77
 physical and chemical constitution, 75
 traditional mind-body problem, 74
 A Weaver's Cottage, 74, 75
- G**
 Global ignition, 112
 Global neuronal workspace. *See* Dehaene's theoretical approach
 The Global workspace theory. *See* Baars's workspace theory
- H**
 Hard problem, 4, 56
The Harp of Erin, 101, 102
 Helen abduction, 118, 119
 Heterophenomenology, 26
 Higher-order perceptions (HOPs), 53
 Higher-order theory of consciousness, 79–83
 Higher-order thoughts (HOTs)
 hard problem, 82
 individual conscious, 80
 introspection phenomenon, 80
 phenomenal consciousness, 82
 reportability, 81
 “significant” and “insignificant” causes, 82
 theory of consciousness, 80
 Holism, 22, 39
 HOTs. *See* Higher-order thoughts (HOTs)
 Human consciousness. *See also* Donald's hypothesis
 and cognition, 117
 development, brain regions, 119, 120
 social dimension and mimetic skills, 119
 transitions, 118
 Humphrey's theory
 evolution of, human intelligence and consciousness, 133
 incommensurable phenomena, 133
 intentional mental states, 134
 “loop” concept, higher-order cognitive process, 136
 perception and sensation, 134
 phenomenological, cognitive and social domains, 136
 properties, 135
 pseudo-sensory phenomenology, 135
 self-portrait, Johannes Gump, 136
 self-resonance/recursive intentionality, 135
 sensory phantasms, 134
- I**
 Incommensurable phenomena, 133
 Inferential role semantics, 39
 Informational atomism, 39
 Inner sense theory
 description, 52–53
 higher-order mental representation, 52
 HOP and HOTs, 52–53
 Rosenthal's model, 53
 Integrated information theory. *See also*
 Tononi's approach
 description, 163
 implications, 166, 167
 quantity and quality, consciousness, 164
 Intentionality
 aspectual shape, 17
 description, 16
 intentional existence/inexistence, 16
 perspectival nature, 17
 psychologism, 15–16
 relational structure, 17
 sensations, 16–17

- “Intentionality is the mark of the mental”, 16
 Intentional mental states, 134
 Intentional stance, 26
 Interactionism, 124
 Interoceptive maps, 106
 Introspective consciousness, 50
 Inverted-spectrum argument, Kim’s, 46, 47
- J**
 Jaynes’ theoretical analysis
 historical texts and archeological data, 139
 human with unconscious minds, 140
 linguistic reentrant processes, 141
 mental events, 140
 metaphor process, 140–141
 narratization, 141
 nervous system implementation, 139
 simple learning tasks, 140
 structions (intentions), 140
 verbal/auditory hallucinations, 143
- K**
 Kim’s philosophy
 analytical argumentation, 43
 causal overdetermination, 45
 conditional thesis, 45
 epiphenomenalism, 45
 inverted spectrum, experiment of, 46, 47
 mental causation, 44, 45
 mind-body supervenience, 44
 principles, 44–45
 The Rainbow Landscape, 46
- L**
 Language of thought hypothesis, 38
 Libet’s theory
 CMF (*see* Conscious mental field (CMF))
 conscious veto awareness, 147
 control function, 147
 criticism, 148
 sensory thresholds, 145
 “Loop” concept, higher-order cognitive process, 136
 Lycan’s philosophy. *See also* Inner sense theory
 evolutionary process, 52
 explanatory gap, 52
 physicalist/materialist approach, 49
 Portrait of a Young Woman, 51
 types, consciousness, 50, 51
- M**
 Main complex system, 164
 McGinn’s philosophy
 acquaintance and knowledge, 56, 57
 mind’s metaphysics and, 55
 neuroscientific theory, 57
 “new mysterianism”, 55
 perspective/paradigm shift, 58
 reductive physicalism, 56, 57
 The Tempest, 58, 59
Meditations on First Philosophy, 31–33
 Mental causation, 34, 44, 45
 Mental events
 anomalous monism, 19, 21
 epiphenomenalism, 21, 45
 intentionality, 16, 22
 “pain”, 74–75
 with physical events, 20, 140
 zombie argument, 5
 Metaphier, 140, 141
 Metaphrand, 140–141
 Microtubules, 159
 Mind-body supervenience, 44
 Morning and Evening Star, 11, 12
 Multiple consciousnesses. *See* Zeki’s theory
 Multiple drafts model
 description, 27
 “fame”, 27
 “final draft”, 27
 global workspace, 27
 homunculus, 26, 27
- N**
 Nagel’s philosophy
 Ariel on a Bat’s Back, 62
 consciousness and subjectivity, 61
 counterintuitive consequences, 63
 human concepts, 63
 Kripke’s analysis, mental properties, 64
 mental reality, 62
 Mind and Cosmos, 64
 The Psychophysical Nexus, 64
 “sufficient similarity”, organisms, 63
 teleological propensities, 65
 truths, 64
 type subjectivism/objective phenomenology, 63
 Naturalistic dualism. *See* Property dualism
 Neural correlates of consciousness (NCCs), 87–88, 100, 101, 103, 175, 178
 Neurobiological framework. *See also* Crick and Koch’s contributions
 of consciousness, 100
 description, 101

- neuronal synchronization, 101–102
 - Neuronal rewiring possibility, 69
 - Neurophilosophy
 - conscious experiences, 10–11
 - decision-making process, 11–12
 - definition, 9
 - discoveries, neuroscientific, 12–13
 - and eliminative materialism, 9, 10
 - Morning and Evening Star, 11, 12
 - Neuropsychology
 - cognitive, isolable functional subsystems, 40
 - Damasio's hypothesis, 105
 - Donald's hypothesis, 117
 - modules, features, 39–40
 - New mysterianism, 55
 - Noë's model
 - The Art of Painting*, 68, 69
 - embodied cognition, 67
 - neuronal rewiring possibility, 69
 - perceptual consciousness, 67–70
 - sensorimotor dependencies, 68, 70
 - visual awareness, 70
- O**
- Objective reduction (OR), 159
 - Order and Chaos*, 165
 - O'Regan philosophy. *See also* Sensorimotor approach
 - analysis, 151
 - eye movements, reading tasks, 151
 - materialist identity theory, 153
 - raw feels/*qualia*, 152
 - Organism consciousness, 50
 - Organizational invariance, 6
 - Origin of the Modern Mind*, 117
- P**
- Pallium* region, 125
 - Panpsychism, 7, 103, 158
 - Paraphiers, 140
 - Paraphrand, 140
 - Passions of the Soul*, 31, 33–34
 - Penrose and Hameroff's theory
 - discrete physical events, 158
 - entity with distinctive qualities, 158
 - general relativity and cosmology, 157
 - objective reduction (OR) and microtubules, 159
 - physicists and neuroscientists, 160
 - property dependent, 158
 - proto-consciousness, 159
 - Penumbra, neuronal activity, 103
 - Perceptual consciousness, 67–70
 - Perceptual projection, 170–172
 - Perceptual visual asynchrony, 175
 - Perspectival nature, intentionality, 17
 - Phenomenal consciousness, 4, 52, 53, 57, 70, 82, 96, 101, 114, 178
 - Phenomenal space concept, 170
 - Phenomenological internalism, 170
 - Philosophical zombie argument, 77
 - Plato's Academy*, marble mosaic panel, 177
 - Plato's Cave*, Flemish School, 130
 - Popper and Eccles philosophy
 - auditory perceptions and feelings, 124
 - causal overdetermination, 126
 - human intelligence and ingenuity, 124
 - interactionism, 124
 - mental and psychological phenomena, 124
 - pallium, 125
 - physical bodies, 124
 - psychons and dendrons, 125
 - scientific and nonscientific theories, 123
 - synaptic transmission, 123
 - Portrait of a Young Woman*, 51
 - Preconscious hypostases, 141
 - Property dualism
 - description, 6
 - double-aspect theory of information, 7
 - organizational invariance, 6
 - structural coherence, 6
 - Propositional attitudes
 - definition, 37
 - folk psychology, 38
 - language of thought hypothesis, 38
 - Proprioceptive maps, 106
 - Proto-consciousness, 159
 - Pseudo-sensory phenomenology, 135
 - Psychologism, 15–16
 - Psychons, 125
 - The Psychophysical Nexus*, 64
 - Putnam's philosophy. *See* Functional isomorphism
- Q**
- Qualia
 - conscious experiences, 129, 130
 - described, 46, 52
 - hard problem, 25–26
 - inversion, 46–47
 - neurobiological framework, 100
 - phenomenal transform, 129
 - properties, 25
 - quale*, 25
 - Tononi *qualia* space, 165
 - Quantum physics. *See* Penrose and Hameroff's theory

R

- The Rainbow Landscape*, 46
 Recursive intentionality. *See* Humphrey's theory
 Reentrant dynamic core, 129
 Reentry process
 binding problem, 128
 description, 128
 Reflexive monism
 cognitive processes, 171
 description, 169
 dualism and reductionism, 169
 mental and physical features, 169
 nonreductionist dual-aspect theory, 169
 phenomenological internalism, 170
 Relational structure, intentionality, 17
 Rosenthal's philosophy
 creature and state consciousness, 79
 HOT (*see* Higher-order thoughts (HOTs))
 mental state awareness, 79–80
 transitive consciousness, 80

S

- Searle's perspective
 causal and ontological reducibility, 86, 87
 "Chinese room" argumentation, 88
 "epistemology", 87
 "intentionality" concept, 85
 "ontology", 86
 Self-consciousness, 50, 100, 115
 Self-portrait
 as autobiographical self, 107
 Johannes Gump, 136
 of young, middle-aged and old
 Rembrandt, 108
 Self-resonance/recursive intentionality, 135
 Sensorimotor approach
 within cognitive self, 154
 description, 151
 insubordinateness and grabbiness
 concept, 154
 phenomenal qualities, 153
 raw feel, 153, 154
 richness and bodiliness concept, 154
 science/robotics, 155
 structure and nature, 153
 Sensorimotor dependencies, 68, 70
 Sensory brain activity, 112
 Sensory phantasms, 134
 Somatic marker hypothesis, 105

- State consciousness, 50, 79
Still Life with a Nautilus Cup, 152
The Storm at Sea (1568), 17
 Structions (intentions), 139–140
 Structural coherence, 6
 Subjective consciousness, 50, 51

T

- The Tao Te Ching*, the opening
 verses of, 88, 89
The Teatro Regio in Turin, 94
The Tempest, Italian Renaissance artist
 Giorgione, 58, 59
 Thalamocortical system
 epiphenomenalism, 130
 functional cluster, consciousness, 129
 limbic value systems, 130–131
 phenomenal transform, 129
 primary and secondary consciousness, 131
 Qualia/conscious experiences, 129–130
 reentrant dynamic core, 129
 Third-person epistemology, 87
 Third-person ontology, 86
 Tononi's approach
 mathematical terms, 164
 qualia space, 165, 166
 sleep mechanisms, 163
 spatiotemporal framework, 164
 Transitive consciousness, 80

U

- Understanding Consciousness*, 169
 Unified field theory, 87

V

- Velmans reflexive theory. *See* Reflexive monism
 Verbal/auditory hallucinations, 143
 Vocabularies, 19–21

W

- The Wanderer Above the Sea of Fog*, 80, 81
A Weaver's Cottage, 74, 75

Y

- Young Slave*, 112, 113

Z

Zeki's theory

binding problem, 176

conscious and unconscious
processing, 176micro-and macro-consciousness,
175–178neural correlates of consciousness
(NCCs), 175

neuroesthetics, 175

perceptual visual asynchrony, 175

Zombie argument, philosophical, 2, 3, 26, 51,
64, 69, 77, 103*Zombie modes*, 102