BLACKWELL READINGS IN PHILOSOPHY

# CONSCIOUSNESS

Edited by

Josh Weisberg

and

David Rosenthal

WILEY Blackwell

Consciousness

## WILEY BLACKWELL READINGS IN PHILOSOPHY

## SERIES EDITOR: STEVEN M. CAHN

*Wiley Blackwell Readings in Philosophy* are concise, chronologically arranged collections of primary readings from classical and contemporary sources. They represent core positions and important developments with respect to key philosophical concepts. Edited and introduced by leading philosophers, these volumes provide valuable resources for teachers and students of philosophy, and for all those interested in gaining a solid understanding of central topics in philosophy.

1 *God and the Problem of Evil*
Edited by William L. Rowe
2 *Epistemology: Internalism and Externalism*
Edited by Hilary Kornblith
3 *Free Will*
Edited by Robert Kane
4 *Democracy*
Edited by David Estlund
5 *Meaning*
Edited by Mark Richard
6 *Philosophical Skepticism*
Edited by Charles Landesman and Roblin Meeks
7 *Consequentialism*
Edited by Stephen Darwall
8 *Contractarianism/Contractualism*
Edited by Stephen Darwall
9 *Deontology*
Edited by Stephen Darwall
10 *Virtue Ethics*
Edited by Stephen Darwall
11 *Personal Identity*
Edited by Raymond Martin and John Barresi
12 *Perception*
Edited by Robert Schwartz
13 *Theories of Truth*
Edited by Frederick F. Schmitt
14 *Social Justice*
Edited by Matthew Clayton and Andrew Williams
15 *Consciousness*
Edited by Josh Weisberg and David Rosenthal
16 *Privacy*
Edited by Steven M. Cahn and Carissa Véliz

# Consciousness

Edited by

*Josh Weisberg and David Rosenthal*

WILEY Blackwell

This edition first published 2022



*Registered Office*
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

*Editorial Office*
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.



*Library of Congress Cataloging-in-Publication Data*

Names: Weisberg, Josh, editor. | Rosenthal, David M. editor.
Title: Consciousness / edited by Josh Weisberg and David Rosenthal.
Other titles: Consciousness (John Wiley & Sons)
Description: Hoboken, NJ : John Wiley & Sons, 2022. | Series: Wiley Blackwell readings in philosophy | Includes bibliographical references and index. | Contents: Introduction / Josh Weisberg and David Rosenthal – What is it like to be a bat? / Thomas Nagel – What is it like to be boring and myopic? / Kathleen Akins – Consciousness and its place in nature / David Chalmers – The explanatory gap / Joseph Levine – A third-person approach to consciousness / Daniel Dennett – What Mary didn't know / Frank Jackson – In defense of the phenomenal concepts strategy / Katalin Balog – What experience teaches / David Lewis – On a confusion about a function of consciousness / Ned Block – The intrinsic quality of experience / Gilbert Harman – How to think about mental qualities / David Rosenthal – Conscious experience / Fred Dretske – The same-order monitoring theory of consciousness / Uriah Kriegel – What kind of awareness is awareness of awareness / Michelle Montague – Higher-order theories of consciousness / Josh Weisberg – Perceptual consciousness as a mental activity / Susanna Schellenberg – The nature of agentive awareness / Myrto Mylopoulos – Realistic monism / Galen Strawson – Property dualism and the merits of solutions to the mind-body problem / Fiona Macpherson.
Identifiers: LCCN 2021037024 (print) | LCCN 2021037025 (ebook) | ISBN 9781119669326 (paperback) | ISBN 9781119669296 (pdf) | ISBN 9781119669340 (epub)
Subjects: LCSH: Consciousness.
Classification: LCC B808.9 .C6664 2022 (print) | LCC B808.9 (ebook) | DDC 126–dc23
LC record available at https://lccn.loc.gov/2021037024
LC ebook record available at https://lccn.loc.gov/2021037025

Cover Design and Image: Wiley

Set in 9/11pt Ehrhardt by Straive, Pondicherry, India

# Contents

# Acknowledgments

The editors and publisher gratefully acknowledge the following permissions granted to reproduce the copyright material in this volume:

Chapter 1
Thomas Nagel, "What Is It Like to Be a Bat?" pp. 435–450 from *The Philosophical Review*, 83:4 (October 1974). Duke University Press, public domain.

Chapter 2
Kathleen Akins, "What Is It Like to Be Boring and Myopic?" pp. 124–160 from Bo Dahlbom (ed.), *Dennett and His Critics: Demystifying Mind*. Oxford: Blackwell, 1993. Reproduced with permission of John Wiley & Sons.

Chapter 3
David Chalmers, "Consciousness and Its Place in Nature," pp. 102–142 from Stephen P. Stich and Ted A. Warfield (eds.), *The Blackwell Guide to Philosophy of Mind*. Oxford: Blackwell, 2003. Reproduced with permission of John Wiley & Sons.

Chapter 4
Joseph Levine, "The Explanatory Gap," pp. 69–92 from Joseph Levine, *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press, 2001. Reproduced with permission of Oxford University Press.

Chapter 5
Daniel Dennett, "A Third-Person Approach to Consciousness," pp. 25–56 from Daniel Dennett, *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press, 2005. Reproduced with permission of MIT Press.

Chapter 6
Frank Jackson, "What Mary Didn't Know," pp. 291–295 from *The Journal of Philosophy*, 83:5 (May 1986). Reproduced with permission of *The Journal of Philosophy*.

Chapter 7
Katalin Balog, "In Defense of the Phenomenal Concepts Strategy," pp. 1–23 from *Philosophy and Phenomenological Research*, 84:1 (January 2012). Reproduced with permission of John Wiley & Sons.

Chapter 8
David Lewis, "What Experience Teaches," pp. 29–57 from *Proceedings of the Russellian Society*, 13 (1988). Reproduced with permission of Cambridge University Press.

Chapter 9
Ned Block, "On a Confusion about a Function of Consciousness," pp. 227–247 from *Behavioral and Brain Sciences*, 18:2 (June 1995). Reproduced with permission of Cambridge University Press.

Chapter 10
Gilbert Harman, "The Intrinsic Quality of Experience," pp. 31–52 from *Philosophical Perspectives*, 4 (1990). Reproduced with permission of John Wiley & Sons.

Chapter 11
David Rosenthal, "How to Think about Mental Qualities," pp. 368–393 from *Philosophical Issues*, 20 (October 2010). Reproduced with permission of John Wiley & Sons.

Chapter 12
Fred Dretske, "Conscious Experience," pp. 263–283 from *Mind*, 102: 406 (April 1993). Reproduced with permission of Oxford University Press.

Chapter 13
Uriah Kriegel, "The Same-Order Monitoring Theory of Consciousness," pp. 143–170 from Uriah Kriegel and Kenneth Williford (eds.), *Self-Representational Approaches to Consciousness*. Cambridge, MA: MIT Press, 2006. Reproduced with permission of MIT Press.

Chapter 14
Michelle Montague, "What Kind of Awareness Is Awareness of Awareness?" pp. 359–380 from *Grazer Philosophische Studien*, 94:3 (2017). Reproduced with permission of Brill Publishers.

Chapter 15
Josh Weisberg, "Higher-Order Theories of Consciousness," pp. 438–457 from Uriah Kriegel (ed.), *The Oxford Handbook of the Philosophy of Consciousness*. Oxford: Oxford University Press, 2020. Reproduced with permission of Oxford University Press.

Chapter 16
Susanna Schellenberg, "Perceptual Consciousness as a Mental Activity," pp. 114–133 from *Noûs*, 53:1 (March 2019). Reproduced with permission of John Wiley & Sons.

Chapter 17
Myrto Mylopoulos, "The Proprietary Nature of Agentive Experience," original in this volume. Reproduced with permission of Myrto Mylopoulos.

Chapter 18
Galen Strawson, "Realistic Monism," pp. 3–31 from *Journal of Consciousness Studies*, 13:10–11 (2006). Reproduced with permission of Imprint Academic.

Chapter 19
Fiona Macpherson, "Property Dualism and the Merits of Solutions to the Mind-Body Problem: A Reply to Strawson," pp. 72–89 from *Journal of Consciousness Studies*, 13:10–11 (2006). Reproduced with permission of Imprint Academic.

# Introduction

Josh Weisberg and David Rosenthal

Right now, you are undergoing the conscious experience of reading this text, combined with a shifting background of sensory, emotional, and cognitive coloring. The conscious experience of the reading, together with the accompanying background feel of sensation, emotion, and thought, make up how things subjectively seem to you, how things appear, as best you can tell, from your own unique point of view. Consciousness is at once acutely familiar—it makes up the experienced moments of your waking (and perhaps your dreaming) life. But consciousness also raises deep and interesting philosophical questions, questions about how any mere physical subject could produce such a wonder, and questions about how there could be a seemingly private and isolated spot of personal subjectivity in an objective, impersonal world. Perhaps the challenge of developing a satisfying theoretical understanding of consciousness is beyond us—we have reached the limits of what we can comprehend. Or maybe today's shortcomings are only temporary barriers to an illuminating theory of consciousness, one properly embedding it in our scientific worldview. And possibly we already have the resources for a satisfactory theory from the way we think about things in commonsense terms.

This reader provides an entry point for considering these and related theoretical questions surrounding consciousness. This introductory section begins with a brief background survey of contemporary debates on consciousness. It then provides a characterization of the notion of consciousness at issue and considers why consciousness understood this way might be theoretically problematic. It follows with a survey of some of the major theoretical positions on consciousness and it closes with a synopsis of the sections of the book.

## I General Background

The contemporary philosophical problem of consciousness has its roots in the traditional mind-body problem, the problem of fitting mentality into the mechanistic, mathematical worldview that emerged with the scientific revolution. Galileo, reflecting on the underpinnings of the new scientific thought, wrote that

> The book [of Nature] is written in mathematical language, and the symbols are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth. (Galilei 1623/1957, 237–238)

*Consciousness*, First Edition. Edited by Josh Weisberg and David Rosenthal.

The great breakthroughs of Galileo and his scientific successors turned in part on the "mathematization" of scientific theorizing. The key to knowing nature, according to modern science, is to capture it in mathematical language, language leading to clear, precise hypotheses to be checked by experiment. Mathematical thinking therefore sits squarely at the heart of modern science. Because of this, anything failing to fit into mathematical terms was in danger of being left out of the scientific story of nature altogether. This was especially pressing when it came to the qualities of conscious experience, like experienced color, sound, and taste. How does one capture basic sensory qualities in language compatible with the Galilean mathematical Book of Nature? Indeed, Galileo himself concluded there were no colors out there in the world at all; rather, color experiences were a reaction we have to the presence of certain mathematically characterizable features in the world, like the reflective surfaces of certain objects. Color and color experience seemed to be cut off from physical reality at the beginning of modern science.

Further, the emerging new science saw nature in mechanical terms. The human body and natural phenomena in general, like the motion of cannonballs and the orbits of the planets, were seen as actions of a great clockwork machine. But if the human body is just a machine governed by physical principles capturable in mathematical terms, as the new science suggests, how are we to account for the mind with its sensory qualities and rational capacities? It is unclear how mechanical theorizing can capture the distinctive qualities of conscious experience and the flexible, creative reasoning of the rational human mind. There seems to be no place for the mind in the theoretical picture of the new science.

René Descartes, often called the father of modern philosophy, developed an influential response to these worries: his famous mind-body dualism. A pressing worry facing the new science was that if our bodies are just machines embedded in a clockwork universe, there seemed to be no room for ideas that are central to the way we think about psychological functioning in commonsense terms, such as the soul and free will. Descartes tried to show that the new science was compatible with our commonsense conception of the mind. He argued that mind and body are fundamentally different substances. Body is extended, nonthinking matter, fully explicable in the terms of the new mechanistic science. But mind is an unextended, thinking substance, one that is not caught in the causal web of mathematical physics. Mechanistic science explains the realm of physical body, but the realm of mentality sits outside this framework, leaving open the possibility of a free and rational mind, able to survive the death of the body. Thus, science and common sense can coexist, on Descartes's theory.

But this leaves the mind outside the physical world. How is it able to connect with the physical body at all? This is known as the "interaction problem" for dualism. Descartes successfully carves off the mind from the clockwork machine, but it is unclear how to reconnect it in everyday life. When a piano drops on my foot, I will likely consciously experience a sharp pain. How does the damage in my physical foot impact my mind, particularly if my mind is unextended, and so takes up no space at all? This worry was pressed on Descartes by Princess Elizabeth of Bohemia. Descartes's answer was straightforward enough: occurrences in the unextended mind causally interact with occurrences in physical reality.

Fair enough; but a problem remains. Over the next few centuries, advances in physics, chemistry, engineering, and other sciences seemed to show that all causation eventually reduces to causation in physics. Chemical interactions can be explained by physics and biological functioning by chemistry. And over the eighteenth and nineteenth centuries, physiology, and eventually the new scientific disciplines of psychology and neuroscience, began to make similar inroads on the realm of the mind, hoping to develop a proper science of the mind. The explanatory successes of the physical sciences led to new attempts to integrate mind into the physical world.

Initial progress in the science of psychology involved the systematic correlation of changes in physical stimuli with changes in psychological reaction. This approach, known as "psychophysics," helped bridge the gap between mind and world, and it is a flourishing branch of psychology to this day. But as psychology moved to investigate more complex "higher" mental phenomena, methodological problems began to appear. Central to the approach of early psychology was the use to detailed introspective reports, reports about what was happening in the minds of subjects as they underwent psychological experiment. But disagreements between subjects about the nature and presence of what was being reported led to intractable problems. If one set of subjects claimed something was present in their experience during an experiment and another set of subjects claimed nothing was, how are we to decide who is correct? There seems to be no public, external check on experiment, a key component of scientific inquiry. The new discipline of psychology was in danger of failing to meet the rigorous standards of science.

In reaction, some psychologists proposed strongly restricting their methodology, developing a "behaviorist" psychology. For these behaviorists, only directly observable phenomena can be studied scientifically. Since we cannot directly observe inner mental states, they cannot be studied in a scientific psychology. However, behavior can be directly observed. So we can base a scientific psychology on observable behavior. In this way, we could avoid the intractable debates which plagued "introspectionist" psychology. Behaviorism of this sort became the dominant view in psychology for much of the first half of the twentieth century.

A parallel move occurred in philosophy around the same time. In the late nineteenth and early twentieth centuries, many in philosophy came to be suspicious of the appeals to our inner life that dominated much of post-Kantian, nineteenth-century philosophical thinking. But our inner psychological lives can be expressed in speech, and that observation led to what is now known, in Richard Rorty's useful term, as the "linguistic turn" in philosophy, a primary focus on the way philosophical issues emerge in the use of language.

Accompanying this shift was a restrictive claim about how much of language worked. Philosophers known as logical empiricists held that only terms have meaning and only if the sentences that they occur in can be verified. If there was no way to verify the sentences that a term occurs in, that term does not mean anything and does not pick anything out. And it was thought that applying this test for meaningfulness, echoing the methodology of David Hume several centuries earlier, could show that a range of philosophical problems are simply meaningless, due to mere confusions of language rather than genuine problems about reality. This approach had an impact on the mind-body problem. If mental terms pick out something private and subjective, we cannot verify if mental terms apply. But it seems that Descartes's view of the mind suffers from this very problem: for him, mental-state terms picked out private, subjective states. So, if Descartes is right about the mind, mental-state terms, like 'pain' or 'belief' ought to be meaningless. But they are not meaningless—we speak meaningfully to each other all the time about pains and beliefs. So Descartes's view of the mind must be wrong. How, then, do our mental-state terms properly function, according to the logical empiricists? The sentences those terms occur in must be subject to verification and, hence, by something observable. And the most likely observable occurrences for that job are type of behavior. That is, the way that words like 'pain' and 'belief' function rests on types of potential behavior that we might observe. Thus, we arrive at an argument for a form of behaviorism from the perspective of language-driven philosophy.

But behaviorism turned out to be overly restrictive in its approach. It suggested that we cannot directly study inner mental episodes scientifically or perhaps even meaningfully talk about them. We might have simply to give up any serious understanding of the very events so central to our mental lives—conscious experiences of sensations, emotions, thoughts, and desires. This prompted a reaction, both in psychology and philosophy, to find a way to speak about and study the mind in a rigorous way, one that both allowed for reference to inner mental states but avoided the epistemological problems of earlier, more unconstrained approaches. The solution was to recognize that in both ordinary talk and scientific theory, we often refer to unobservable *posits*, picked out by terms we introduce theoretically to explain the things we can observe. A prime example is the term 'electron', which picks out a posited subatomic particle. We cannot directly observe electrons, but they allow us to successfully predict and explain what happens in a wide range of observable situations. This gives us good reason to believe electrons exist. Further, ordinary language functions just fine when employing this sort of device. The antidote to behaviorism's narrow vision is to allow that mental states are *posits* in a theory, states posited to explain observable behavior. But what sorts of states?

In parallel to these developments in philosophy, results in psychology showed that behaviorism failed to explain certain sorts of mental functioning. Animals possess inner structures allowing them to navigate the world beyond the simple stimulus-response connections allowed in behaviorism. These "cognitive maps" are rich inner structures underwriting an animal's ability to negotiate complex environments. Further, the mental resources needed to acquire language seemed beyond the reach of simple behaviorist theorizing. To fill the gap in theorizing, advances in computer science were introduced into psychological and linguistic theorizing. Alan Turing and others developed the foundations of what would become known as the "computational theory of mind," the idea that mental states are posited inner states computationally mediating between perceptual inputs and behavioral outputs. They play their role by computing what the organism should do next given its current input and goals and activating those responses. On this theory, mental states are computational states. The approach had wide application in what would become known as "cognitive science," the scientific

successor to behaviorism. Further, it fit well with the view of using theoretical posits to expand the conditions for acceptable scientific theorizing and everyday language use. Mental states, on this view, are posits of a theory, and they are theorized to be computational states. Computational states, in turn, are defined by their connections to observable perceptual input and behavioral output—to observable phenomena. So they are not essentially hidden and private.

The computational theory of mind defines mental states in terms of what they do, in terms of the *functional role* they play. The theory has proven successful at providing at least a preliminary explanation of a range of complex mental behavior, including rational inference, learning, memory, and other processes. This "functionalist" approach stands as the received view in theorizing about the mind in contemporary philosophy and psychology, though there are many unsettled issues. And it is here that the major contemporary debates about consciousness, those addressed in this reader, begin. With the functionalist computational theory of mind, we have the outlines of a promising explanatory story about how much of mentality can fit into the picture of modern science. But when we focus on consciousness, there still seems to be something missing. A number of philosophers argue that although progress has been made with the mind in general, consciousness still remains outside of our scientific understanding. It is a residual element of the mind-body problem, the last bit seemingly resisting explanation. But consciousness is central to who we are subjectively, so we are left with a philosophical puzzle. But can we be more specific about what we mean by 'consciousness'? And why think that consciousness remains left out from our scientific worldview?

## II The Study of Consciousness

We use the term 'consciousness' in a number of ways in ordinary speech. One way is to distinguish conscious from unconscious creatures. If a creature is active and responsive to its environment, we consider it to be conscious in this sense. If it is incapacitated and unresponsive to its environment, we consider it to be unconscious. We can call this idea "creature consciousness," as it pertains to the condition of a person or other creature. This is not the notion of consciousness leading to the philosophical worries here. Creature consciousness is plausibly a phenomenon explicable in biological terms, in terms of the proper biological functioning of the organism in question. While there is certainly a great amount of biological and physiological complexity at issue here, it is not especially mysterious how a creature could be conscious in this sense, given the explanatory resources of biological science.

However, we also sometimes use the term 'consciousness' to apply to mental states. We speak of consciously seeing a friend in a crowd, or consciously hearing the key change in a piece of music. Or we may become conscious of our lingering guilt over eating the last cookie in the jar. In such cases, a mental state—a state of seeing, hearing, or feeling an emotion—is conscious, as opposed to being nonconscious. There is a special sort of difference when our lingering guilt goes from an underlying nonconscious state to a conscious state. We are now aware of our guilt in way we were not just before. Likewise, I may be conscious of seeing the crowd but not conscious of seeing my friend. Then I consciously see her—I become aware of her in a conscious way. Common sense and psychological science both accept that mental states can occur consciously or nonconsciously. The kind of consciousness at issue here we can call "state consciousness." Often, our mental states occur nonconsciously, but sometimes, they occur consciously. When they do, there is something it is like for us as subjects to be in those states—there is something it is like to be us, *for* us, in Thomas Nagel's terms (see Chapter 1). State consciousness is the phenomenon raising the philosophical questions that this reader is most concerned with.

What, then, is state consciousness? The best way to get a handle on any term is to see what it contrasts with. So we can zero in on the idea by focusing on the contrast between conscious and nonconscious mental states. Sometimes, we have the feeling of knowing the name of a person but cannot bring it to mind. The information remains outside of our conscious awareness. But then the name comes to us. And then the specific knowledge becomes conscious—our state of knowing that friend's name becomes a conscious state. More dramatically, you may have had the experience of walking across campus deep in thought, only to look up and realize you have arrived at your classroom or the library. But you may have no recollection of what you saw on the walk. Those states of seeing were plausibly nonconscious, as your mental focus was elsewhere. Still, you did not crash into anything, nor did you trip and fall. Further, you arrived at the right place, all indicating that nonconscious visual states guided your actions. But if a fox had suddenly leaped into your path, your visual state would have become conscious—you would have consciously seen

the fox. This transition, from not present to us to being present to us—*for* us—is the transition from a state being nonconscious to a state being conscious. Further, consider how your elbow (or big toe or belly button) feels right now. You likely were not conscious of those feelings prior to this prompt, but now you are. This, again, reasonably marks the transition from nonconscious to conscious state. These examples are familiar and everyday. Theorists differ over how to best cash out this commonsense distinction between conscious and nonconscious states, as the following chapters will make clear. But it gives us a good starting for thinking about consciousness.

What is more, the difference between conscious and nonconscious mental states has been widely studied in psychology and neuroscience. In "priming" studies, stimuli are flashed at subjects so quickly that subjects report not seeing anything. But there is reliable evidence that the flashed "prime" influences subsequent behavior, despite the fact that it remains nonconscious. Subjects can nonconsciously process the meaning of words, complex pictorial scenes, and even the emotional impact of a stimulus. All this occurs without the subject consciously seeing what is influencing them. But if the same stimuli are presented slowly, subjects consciously see them and can report doing so. The difference in the fast and slow cases, from the subject's point of view, marks the boundary between conscious and nonconscious states.

We see the same sorts of things in more unusual neurological cases. Subjects with brain injury sometimes lose the ability to consciously see things on one side of their visual field. But information presented in the "neglected" area can still influence their behavior in complex ways. And some subjects with damage to the visual areas of their brains have large "blind fields" in their visual perception. Still, they can, employing what is known as "blindsight," guess correctly at a high rate about what is present in the locations they cannot consciously see, indicating that complex visual information is being processed and registered. These cases from psychology and neuroscience highlight that mental states, including states of visual perception, can occur nonconsciously as well as consciously. Ordinary and scientific understanding both mark this difference. The central challenge of consciousness is to explain the nature of this difference, to capture and illuminate what is special about this transition from the unnoticed darkness of nonconscious mentality to the present, lived reality of conscious experience.

But why think this explanatory challenge poses any special type of problem? One worry, stressed by Descartes, has to do with the presence of a first-person subject in consciousness: our conscious states are experienced as fundamentally our own. Indeed, there is a sense in which we subjectively just are our stream of consciousness—we may seem to ourselves just to be this particular procession of conscious states. It is the unique perspective we have on the world, our very own subjective point of view. But scientific explanation aims for an objective picture of the world, a "view from nowhere," as Thomas Nagel puts it. How can an objective scientific worldview capture the subjectivity of conscious points of view? How does this sort of subjective perspective emerge from objectively characterized matter? Some contend that this is merely a special, but tractable, engineering problem, a puzzle of biology and neuroscience, but not one requiring great speculative leaps (see Akins, Chapter 2). But others, following Nagel, worry that the gap between objective and subjective is too broad to bridge by ordinary scientific means. We may be in the presence of something unique and different, and something fundamental to who we are. Subjectivity is one of the key problems prompted by considering consciousness.

Another central worry about consciousness involves the distinctive qualities of conscious sensory experience. This is the worry brought on by the Galilean mathematical approach of modern science discussed above. When we consciously see a sunset or consciously hear a jazz trio, we have experiences marked by distinctive sensory qualities—the way things consciously look or sound to us. The reds, yellows, oranges, and grays of a deepening sunset or the subtle timbre of piano chords, plucked bass notes, and plinking cymbals are present to us in conscious experience. They make up "what it is like for us" in such moments. But sensory qualities have long marked off a problematic break in thinking about the natural world. To reiterate, Galileo's mathematical theorizing works well for many features of reality and has led to the great scientific breakthroughs. But it leaves the sensory qualities in a difficult position. How can they be captured in such mathematical-geometric language? Further, sensory qualities as we consciously access them seem simple and lacking in structure. They appear to be the basic building blocks of experience, not decomposable into anything more basic. Red, orange, sweet, sour, loud, soft—there seems to be little one can say to explain the "conscious feel" of such things if another has not experienced them. We have reached the explanatory

bedrock of the mind, it seems. It may be, however, that there are ways to decompose and integrate sensory qualities into our scientific worldview without radical revision. Or perhaps we have reached the limits of our ordinary understanding and radical measures are called for. Controversies over sensory quality are central to the philosophical issues surrounding consciousness.

Yet another issue raised by consciousness is its apparent *unity*. We seem to ourselves to be seamless and complete, to be single unified subjects, rather than a conglomeration of parts, however well those parts may work together. Consciousness appears unified along two dimensions: at one time, consciousness is bounded and self-contained—a single "me" has the experiences. And over time, my conscious experiences seem to happen to the same person, to one temporally extended individual. In the first case, my conscious sensations, the way things look and feel to me, as well as my occurrent conscious thoughts all seem bound together into a continuous stream. How could disparate neural mechanisms issue in this sort of unity? And over time, my experiences all happen from the same subjective point of view, to the same person. This deep philosophical problem of personal identity, of saying what makes me *me* over time, is wrapped up with consciousness. One plausible answer to this question is that it is one and the same consciousness experiencing my life over time. But what makes this consciousness the same as the one that I had as a child or even yesterday? Since brain cells die off over time and the chemical makeup of our nervous system is constantly changing and adapting, it might seem that no underlying physical story could account for the unity of consciousness over time.

As suggested above in considering the development of the functionalist computational theory of mind, many of these concerns emerge because of the problem that consciousness appears to resist a *causal account*. A causal analysis analyzes a concept in terms of what its referent does, in terms of the role the thing picked out by the concept plays in making a causal system function properly. A mouse trap, for example, is a device that takes live mice as input and delivers nonliving mice as output. Do that job, and you are a mousetrap, be you a spring-loaded device of metal and wood or a cat. This provides a causal analysis of the concept "mouse trap"—now we know, in causal-functional terms, what the concept picks out. Likewise, to count as the psychological process of learning, a process must correctly alter a system's reactions to incoming stimuli in light of past stimuli. If I go down a path and stub my toe on a rock, and then next time down the path I take appropriate evasive action, I have learned, and the system in me achieving this result is a learning system. Learning, like trapping mice, is a matter of doing the right thing—of instantiating the right sort of causal process. If things unfold causally in this way, learning has occurred, whatever the underlying nature of the system involved. "Learning" therefore can be causally analyzed.

Indeed, it may be that most concepts used in characterizing the natural world can be analyzed in this way. And this has been a boon to scientific explanation. When we have a causal analysis of a concept picking out a phenomenon, there is a clear recipe for fitting it into that phenomenon worldview of physics: we show how such a causal process is *realized* by the presence of physical parts, properly connected. For example, we can characterize genes as those things carrying hereditary information from parent to offspring. If something does this job (refined in the ways dictated by modern genetics), it is a gene. The discovery that DNA plays this very role allows us to "locate" the gene in a physicalist worldview: it is just this bit of chemical machinery doing the right thing. And the chemistry, in turn, can be unpacked in atomic and then subatomic terms. There is no remaining philosophical conundrum about how genes could be physical (though there remain all sorts of "easy" problems in genetics and microbiology!). Causal analyses allow us to fit big, complex things into a world ultimately made up of small and simple fundamental physical particles.

Still, some authors continue to have the concern that consciousness may fail to be fully analyzable in this way. There may be more to consciousness than its causal profile. It seems that for any causal role we can think of consciousness as playing, we can easily imagine that same role being played without consciousness. In psychology, we specify the causal roles that a mental state plays by spelling out the inputs causing the state to occur, the outputs caused by the state, and the causal interactions that the state has with other mental states. These input, output, and interactive specifications uniquely pin down the mental state in causal terms. Consider, for example, a state of pain. It is caused by a range of inputs: having a piano dropped on your foot, a paper cut, a bee sting, and so on. It causes a range of reactions: screaming, writhing, action to get away from the source of the trouble. And it interacts with various other mental states: it makes it harder to do mathematics, it causes the desire for the pain to stop, it shifts one's attention to the pain. And so on. It will no doubt be a very long and complex story.

But we can in principle spell out pain's causal role in this way.

But what about the *feel* of pain, its distinctive unpleasant "ouchiness"? Can this be fully captured in a causal story? It seems that we can imagine the very same causal processes going on—the same inputs, outputs, and interactions—without any ouchiness at all. All might go on "in the dark," on autopilot, without any accompanying feel. Or so it seems. David Chalmers (Chapter 3) spells out this worry by positing the conceivability of what he calls "zombies": creatures physically just like us, down to the last quark, but lacking in consciousness. Zombies have states playing all the same causal roles as we do, but those states lack the distinctive felt qualities of consciousness, the ouchiness of pain, the tasted sweetness of sugar, the experienced reds and oranges of a sunset. Chalmers contends that the fact that we can easily conceive of such beings shows that there is more to consciousness than the causal role it plays in our psychology. If consciousness were exhaustively analyzable by its causal role (the way a mouse trap is), then we would not be able to imagine zombies. Can you imagine something that reliably takes live mice as input and delivers dead mice as output that is nonetheless not a mousetrap? It may of course be other things—a paper weight, the family pet—but it will certainly be a mousetrap. But something might have states that play the causal roles in question and still not be conscious. Or so Chalmers's case seems to show.

Others, however, contest this claim (see, for example, Dennett, Chapter 5), but we have at least a prima facie worry here. There is a difficulty concerning the epistemology of consciousness—how we know about it. It appears that its presence cannot be guaranteed by looking at the causal behavior of the organism. And it might be that if something's presence cannot be guaranteed in this way, this shows something about what it is made of, something about the metaphysics of consciousness. Physical things generally can be pinned down by their causal profile. If consciousness cannot, perhaps it is not something physical. Or perhaps we need to expand our view of the physical to include consciousness (see Strawson, Chapter 18).

Finally, there is a worry about how we might informatively communicate about consciousness at all. If consciousness is subjective in a special way, and if it defies analysis in terms of its causal role, it is not clear how we could informatively describe and explain consciousness to someone who has not experienced it themselves. This seems to apply most clearly again to the distinctive qualities of consciousness—the ouchiness of pain, the experienced colors of a sunset—but it also makes trouble for describing a unique, subjective point of view. And because there seems to be no such difficulty informatively communicating about physical things, this again suggests that consciousness might not be physical at all. Consider that if someone has never tasted chocolate or seen a sunset, there is a limit to how much we can do to convey these things to them. We can use various comparative terms, but at some point, we run out of things to say beyond, "You have to taste it or see it for yourself!" Frank Jackson (Chapter 6) leverages this worry into an argument against the idea that consciousness is a physical phenomenon. He asks us to consider a super scientist of the future who knows all that a completed physical science can tell. But she has been raised in a special black, white, and gray environment—she has never seen colors like red or green.

Jackson asks us to consider what happens when this scientist is released from her impoverished room and sees red for the first time. Does she learn something new? Jackson contends that she does: she learns what it is like to see red. But she already knew all the facts of physical science. So there must be facts beyond what physical science can tell us. And these are facts about conscious experience, about what an experience of red is like subjectively, "from the inside." Again, there are a range of replies to this challenge (see Balog, Chapter 7, and Lewis, Chapter 8), but there appears to be something unique and different about consciousness, something perhaps indicating that it lies outside of our scientific worldview or stands as a special nonphysical element sitting at the core of our being. Or maybe it just seems to be this way, despite being fully explicable in scientific terms and fitting without remainder into the physical world. In any event, there is a rich space of philosophical debate here, one touching on a variety of deep questions about our place in the natural world and the limits of our understanding.

## III Theoretical Responses to the Problem of Consciousness

In the face of these puzzles, a wide range of responses has been developed, from denying that there is a real philosophical problem at all to making radical revisions to our metaphysical theories of the world. Some philosophers argue that once we find the right way to characterize consciousness, it becomes apparent that there is

nothing deeply problematic here—it only seemed that there was. To the point, perhaps consciousness is a fully functional process, one completely analyzable in causal-functional terms, even though it does not seem to be from the first-person perspective. Consciousness is really just like other complex physical phenomena, but our subjective access to it obscures this fact. Further, it is often the case that advances in science override our pre-theoretic intuitions about things. Given the success of relativity theory and quantum mechanics, we have come to realize that space and time are not what they seemed to naïve common sense. Perhaps the same thing will occur with the successes of psychology and neuroscience. We will come to see that consciousness is physical after all, despite our gut feelings (or what are sometimes called philosophical intuitions) to the contrary. This sort of approach to consciousness—that it is really just a normal, if complex, physical phenomenon—is developed in Chapter 2 by Kathleen Akins, in Chapter 5 by Daniel Dennett, and in Chapter 8 by David Lewis, among others.

Another response to the problem of consciousness is to hold that while consciousness cannot be given a causal-functional analysis, it still can be identified with physical processes in the brain. Consciousness may be something special and different in that it is not *explicable* in more basic physical terms. Still, we might find that consciousness only occurs when certain physical processes are present and then identify consciousness with these processes. There is no further explanation of *why* such processes should result in conscious experience, but it simplifies our worldview to hold that there is just one thing here (a conscious physical process), rather than two things (consciousness and a physical process occurring in the same time and place). This sort of claim is not unprecedented in science. Sometimes, we simply identify one thing with another, as we did with the discovery that electricity and magnetism are one unified force. But it is unusual to do so with complex, macro-level processes like consciousness. Still, the unique appearances of consciousness as subjectively accessed may justify the identity claim. Ned Block takes up this sort of approach in Chapter 9, and it is discussed by David Chalmers (in Chapter 3) and Fiona Macpherson (in Chapter 19) as well.

But some have contended that the proper reaction to the philosophical challenge is to hold that consciousness is something utterly different and distinct from physical stuff. They embrace modern-day versions of Descartes's famous mind-body *dualism*. Some hold that the physical story of humans is incomplete. It requires the addition of nonphysical properties to account for conscious experience. These properties, known as "qualia," make up what it is like for the subject in conscious experience. They are the experienced redness of red or the felt painfulness of a conscious pain. Dualists argue that there could in principle be creatures physically just like us but lacking in consciousness. Consciousness in us is accounted for by the presence of these additional nonphysical properties. But how do these nonphysical properties ever interact with our physical bodies? Some dualists hold that there are special "psycho-physical" laws linking the conscious properties and our physical bodies. In addition to the basic laws of physics, we need to add in these laws, linking conscious properties to nonconscious properties, our physical bodies. This is indeed a radical revision of our scientific worldview, but dualists argue that the change is warranted by the problems presented by the conscious mind.

We are left, however, with the difficulty of explaining how nonphysical conscious properties interact with the physical body. This problem is made more pressing by the widely accepted idea that the physical world is "causally closed," that all physical events have fully physical causes, without the addition of any further influence. This conservation principle has been helpful in developing and expanding the physical sciences, but it looks incompatible with the idea that nonphysical conscious properties can have a causal impact. Some dualists reject the causal closure of physics in order to causally connect the conscious mind to the physical body. And it may be that the probabilistic theoretical picture offered by quantum mechanics opens up a small space for the causal influence of consciousness. Or the staggering causal complexity of the brain may leave open the possibility of outside causal effects, as of yet unnoticed in brain science. Still, these theoretical approaches are in tension with the success of causal closure. An alternative is to hold that consciousness is causally impotent—it is synchronized with the brain, but it has no causal impact of its own. This view, known as epiphenomenalism, may appear counterintuitive (it sure *seems* like our conscious thoughts and feelings cause us to act), but, again, in the face of scientific progress we may have to give up on our pre-theoretic intuitions, no matter how deeply held. In the final analysis, contemporary dualists would rather deal with the difficulties of the interaction problem than reduce or identify consciousness with something merely physical. These dualist positions are discussed in

Chalmers's and Jackson's chapters (Chapters 3 and 6, respectively).

Another response to the problems of consciousness agrees with the dualist that there is something about conscious experience that is not explicable in terms of the ordinary physical properties and processes. But that does not mean we must posit nonphysical properties separate from (though somehow attached to) physical reality. Rather, we can focus on an apparent gap in physical theory and "locate" conscious properties there. Physics arguably deals in relational facts, facts that can be specified by seeing how various properties and processes relate and interact with one another. An electron, on this view, is fully characterized by the relations into which it enters. It is attracted to particles of opposite charge, repelled by those of same charge, and so on. If you enter into these relations—if you play this causal role in terms used above to describe functionalism—you are an electron.

But what is the thing itself that plays this role, the thing that enters into these relations and connections? Physics has nothing more to say here. It does not say anything about the intrinsic or nonrelational properties of electrons, beyond that they must somehow allow electrons to do what they do. But this opens a space to posit that among the nonrelational, intrinsic properties are "phenomenal properties," the properties constituting the felt qualities of conscious experience. Physics does not rule out the presence of such properties—it is silent about the nonrelational nature of physical matter. So perhaps at the core of fundamental physical matter lies a spark of experience, available to bring our consciousness into existence when activated or combined in the right way. It is a matter of debate among theorists pursuing this approach whether these intrinsic phenomenal properties are best seen as physical or nonphysical. Galen Strawson, in his contribution in Chapter 18, argues that they are physical. Chalmers in Chapter 3 surveys other possibilities. This sort of view, known as "panpsychism" or "neutral monism," has the surprising result that there is a bit of consciousness in all things, right down to the smallest subatomic particle. The challenge of explaining how such phenomenal elements scale up to form our experience, as well as the counterintuitive nature of the claim that all things are to some degree conscious, are difficulties facing this theoretical approach. For criticisms, see Chapter 19 by Macpherson.

Taken together, we see a range of deep questions about the nature of consciousness and how we know about it. And we have a variety of interesting and challenging responses to these questions. Consciousness may mark the boundary of our scientific knowledge. It is also central to our concept of ourselves, to who we are as living, thinking beings. And its study provides an excellent introduction to issues in contemporary epistemology, metaphysics, and the philosophies of mind and language. The final section of this introduction provides a brief summary of each chapter.

## IV Chapter Summaries

### *Part I: Problems of Consciousness*

#### *Chapter 1: "What Is It Like to Be a Bat?" by Thomas Nagel*

Nagel contends that "consciousness is what makes the mind-body problem really intractable." Nagel argues that no causal-functional analysis can fully capture consciousness. Further, central facts about consciousness—facts about what it is like to be the organism, *for* the organism—are "subjective facts," facts only knowable from one type of point of view. If my type of point of view is too different from the type of point of view of the conscious subject, I will not be able to know those subjective facts. And since science aims to put things in objective terms, subjective facts will always be something left out of a scientific theory of consciousness. Nagel supports his claim by reflecting on the possible subjective experience of a bat as it echolocates. It seems there is a factual question about what it is like to be the bat, for the bat. But our subjective point of view is too distant from that of bats, so we cannot know this subjective fact. And since subjective facts cannot be given in objective terms, we cannot incorporate this sort of knowledge into science. We are left with a puzzle, one that is at the heart of contemporary philosophical debates about consciousness.

#### *Chapter 2: "What Is It Like to Be Boring and Myopic?" by Kathleen Akins*

Akins provides a response to Thomas Nagel's worries in his "What Is It Like to Be a Bat?" Akins notes that Nagel sees objective science as a "view from nowhere" that will inevitably leave out the subjective "what it's like for the subject" features of experience. To address this worry, Akins reviews a wide range of what is scientifically known about what goes on in bats when they

hunt by echolocation. She shows that we can reasonably infer many facts about the bat's point of view, facts Nagel suggested would be left out of a scientific story. But we find, according to Akins, that there is little to indicate that bats develop a perceptual point of view involving a three-dimensional map of the objects and features of their environments. Thus, it is not clear that we would consider them as having a point of view at all in the sense we usually mean. But, importantly, the remaining outstanding questions about a bat's point of view plausibly involve understanding how they represent their environments and their bodies, questions still within the scope of scientific investigation. Thus, it is premature of Nagel to claim we cannot know what it is like to be a bat.

### *Chapter 3: "Consciousness and Its Place in Nature," by David Chalmers*

Chalmers summarize the major arguments for there being a substantial problem of consciousness, one that cannot be solved by the ordinary means of modern science. He terms this the "hard problem" of consciousness, to distinguish it from the "easy" problems solvable by the ordinary causal-functional approach of science. Chalmers then canvases three major arguments against a materialist theory of consciousness, one holding that consciousness is an aspect of the material world—the explanatory argument, the conceivability argument, and the knowledge argument. Chalmers then delineates a range of responses to these problems, providing an alphabetic categorization of approaches to consciousness. Type-A materialism holds that despite the apparent problems, consciousness can be fully explained in material terms. Type-B materialism holds that full explanation in material terms may be beyond us, but we still have good reason to identify consciousness with material processes. Type-C materialism holds that our current lack of understanding of consciousness does not mean consciousness is not fully material. The remaining responses hold that we must abandon materialism for something different. Type-D and type-E dualism hold that conscious properties are nonmaterial. Type-D contends that these nonmaterial properties can still be causally efficacious, while type-E dualism embraces epiphenomenalism. Finally, type-F theories posit a further element to the physical world, one underlying or constituting basic physical matter.

### *Chapter 4: "The Explanatory Gap," by Joseph Levine*

Levine argues that any attempt to explain consciousness in physical terms will always be left with an "explanatory gap," a range of unanswered explanatory questions that a successful explanation ought to close off. Levine defends a view of explanation taken from the philosophy of science, arguing that in a good explanation the claims doing the explaining must entail the claim being explained. With this conception in hand, he contends that unlike the case of water being explained as $H_20$, the claim that specific brain states are conscious states fails to provide the needed entailment. This, Levine argues, is because the identity of brain states and conscious states is "gappy": it still allows for intelligible requests for explanation, requests that are not intelligible in cases of non-gappy identities like water = $H_20$. Levine considers a number of responses to this problem of the "explanatory gap," and maintains that it presents a serious challenge to a theory of consciousness, even if we have well-motivated reasons to accept that consciousness is ultimately a physical process.

### *Chapter 5: "A Third-Person Approach to Consciousness," by Daniel Dennett*

Dennett challenges the idea championed by Nagel and others that consciousness cannot be studied scientifically due to its private, subjective nature. Dennett argues that these claims of scientific intractability illicitly assume that first-person reflection on consciousness provides infallible direct evidence of something beyond the reach of science. But this begs the question against a workable science of consciousness. Why should we assume that our first-person access is infallible in this way? Instead, Dennett proposes a neutral method of fixing the data that a theory of consciousness must explain: the "heterophenomenological method." Heterophenomenology collects the first-person reports of subjects about consciousness and catalogs the beliefs that these reports express. This is data to be explained. It is not given that subjects are always correct about what they believe concerning consciousness. Though they may believe consciousness is private and nonphysical, this does not mean that it is so. Rather, science is charged with investigating the origin of these beliefs. It may be that science will fail to explain the origin of these beliefs, vindicating those like Nagel. But it may be that science comes to explain everything required. In any event, we cannot just assume

at the outset of investigation that consciousness is beyond scientific explanation.

*Part II: Consciousness and Knowledge*

### *Chapter 6: "What Mary Didn't Know," by Frank Jackson*

Jackson presents his "knowledge argument" against physicalism. Physicalism is the claim that everything is fully physical in nature, consciousness included. If physicalism is true, all the facts are just physical facts. Jackson asks us to consider a super scientist of the future named Mary who knows all the physical facts of a completed science. However, she has been raised in a special black, white, and gray environment—she has never seen color. Jackson asks what will happen when she is released from her black and white confinement and sees red for the first time. Will she learn anything new? Jackson contends that if physicalism is true, she should not, as she already knows all the physical facts. But, Jackson asserts, she will learn *what it is like for her to see red*. This fact about conscious experience was not contained in the physical facts, so physicalism is false.

### *Chapter 7: "In Defense of the Phenomenal Concepts Strategy," by Katalin Balog*

Balog presents a defense of the idea that our knowledge of conscious sensory quality is mediated by special "phenomenal concepts," concepts using sensory qualities themselves to present the qualities. On this model, our first-person access to pain or a sensation of redness involves painfulness or redness actually being instantiated in the thought that we are in pain or seeing red. The phenomenal concepts strategy allows for a defense of physicalism. The epistemic differences between our first-person acquaintance with consciousness and its physical description are explained by this difference in conceptual access. When we think of pain from the first-person perspective, we do so by feeling pain and "pointing" to it mentally. When we think of pain in a physical theory of the mind, we access pain by reading a complex theoretical description of it. This difference misleads us into thinking we are in the presence of two distinct kinds of properties. But according to the phenomenal concepts strategy we access one and the same physical thing, a brain state, by way of two distinct conceptual pathways. There is a dualism of concepts, not a dualism of properties. Further, because of this difference in conceptual access, there is no a priori entailment from the physical facts to the phenomenal facts. Still, that does not mean that consciousness is nonphysical—we can have good a posteriori reasons to posit an identity between conscious states and physical states.

### *Chapter 8: "What Experience Teaches," by David Lewis*

Lewis considers what might follow from the claim that "phenomenal facts"—facts about what it is like to be a conscious organism *for* the conscious organism—can only be gained by having the right kinds of experiences. Jackson contends that this shows that there is a special sort of phenomenal information and that this information cannot be physical information. Lewis allows that such information, if it existed, would indeed be a problem for physicalism, and, further, a number of physicalist responses to the problem miss the point. But Lewis maintains that there is no good reason to believe in phenomenal information. Rather, the knowledge gained about consciousness by way of experience is not factual knowledge at all. It is instead a kind of "know-how"—an ability learned by experience, not a set of facts about experience. All the facts about experience are indeed physical facts, but not all of the abilities taught by experience can be reached by "book learnin'." Physicalism does not claim that all know-how can be gained by reading a description of the physical processes instantiating the ability, so the knowledge argument against physicalism fails.

*Part III: Qualitative Consciousness*

### *Chapter 9: "On a Confusion about a Function of Consciousness," by Ned Block*

Block argues that the use of the term 'consciousness' is ambiguous among several meanings. One is a functional notion that Block labels "access consciousness." Access consciousness involves the global sharing or broadcasting of information around the mind and it is associated with verbal reports, reasoning, and the rational guidance of action. Notably, it is a functional notion, one explicable in causal, functional, or intentional terms. But there is another notion of consciousness, one picking out experiential states in terms of what it is like for the subject to have them. This idea, which Block labels "phenomenal consciousness," is not, he contends, analyzable in causal, functional, or intentional terms. Further, it is

the notion causing the theoretical worries about consciousness. Paradigm cases of phenomenal consciousness involve states of qualitative consciousness, like the conscious experience of pain or the conscious visual experience of red. A central confusion in the study of consciousness, according to Block, is thinking that by explaining access consciousness, one has thereby also shed light on phenomenal consciousness. This equivocation leads to many failed attempts in cognitive science to explain the mystery of consciousness.

### *Chapter 10: "The Intrinsic Quality of Experience," by Gilbert Harman*

Harman defends the thesis of "representationalism": the idea that all the features of conscious experience are ultimately representational in nature and can be explained by an appropriate theory of mental representation. This stands in contrast to those theories holding that qualitative consciousness has an intrinsic nonrepresentational qualitative core. Harman argues that consciousness is fully representational. Every element in experience presents the world or our bodies as being some way, as possessing this or that property. When we consciously see a tree, we are aware of the properties that the tree presents to us, its brown bark and green leaves, and so on. According to Harman, we are never aware of any further qualities of experience alone, qualities not ascribable to things we represent in our world. But if all the properties in conscious experience are representational, a theory of representation should fully explain conscious experience, even experiences of pain or redness. Harman then argues that we have good candidates for naturalized theories of representation, theories fully explicable in physical terms. Therefore, we have a promising route to fully explaining consciousness in terms of representation.

### *Chapter 11: "How to Think about Mental Qualities," by David Rosenthal*

Rosenthal argues that any apparent intuition that the nature of mental sensory quality is accessible only from the first-person perspective is the product of theorizing, rather than something given in introspection. With this in mind, he develops an account of sensory qualities based solely on third-person considerations. That in turn results in an informative theory of qualitative consciousness, a possibility that the rival first-person-only approach rules out from the start. Central to this proposal is the idea that qualitative mental states are not essentially conscious. This is borne out in both empirical results and commonsense usage. Rosenthal's approach to mental qualities—the "quality space" theory—characterizes the qualities in terms of their role in perception. We are able to perceptually distinguish a range of similarities and differences in perceived stimuli. We can use these patterns of discriminative ability to build a space of perceptible stimuli, and infer from that a parallel internal "space" of qualitative mental properties that enable us to make these discriminations. This approach captures all the fine-grained distinctions available to consciousness, but it rests just on considerations available from the third-person perspective. When combined with a plausible naturalist theory of consciousness, for example, Rosenthal's higher-order theory of consciousness, we arrive at a promising route to a satisfying and informative explanation of qualitative consciousness.

## *Part IV: Theories of Consciousness*

### *Chapter 12: "Conscious Experience," by Fred Dretske*

Dretske argues for a distinction between "fact awareness" and "thing awareness." In fact awareness, we are aware that some $x$ is $F$—that I am seeing an armadillo, say. But in thing awareness, we can be aware of the armadillo without being aware that it is an armadillo, or that it is even an animal. It is a more basic kind of connection to the object of our awareness. With this distinction in hand, Dretske contends that mental states can be conscious even if we are not conscious of those states and he offers an example where one sees two different stimuli without being conscious of a difference between them. But such a difference is plausibly seen, so we must be thing aware of the difference, according to Dretske. Therefore, we can be conscious of things by being thing aware of them, but without being fact aware that we are in a state of consciousness. This runs against another class of theory, the so-called "higher-order" theories of consciousness that hold that a mental state is conscious when we are conscious of ourselves as being in it. Dretske instead argues that a mental state is conscious when it makes us aware of the world and it plays the role of being available for the formation of perceptual belief. This "first-order" theory rejects any inner spotlight of consciousness. Instead, consciousness is a matter of having the right sort of representational states playing the proper roles.

### *Chapter 13: "The Same-Order Monitoring Theory of Consciousness," by Uriah Kriegel*

Kriegel introduces and defends the "same-order monitoring" theory of consciousness. On this view, a mental state is conscious when it represents itself as well as the world. There are a number of ways to flesh out this proposal, and Kriegel runs through 11 permutations of the view, in order to establish the position's strengths and to make clear the various theoretical alternatives. One of the central proposals developed involves a multipart mental state with a part representing other parts in an integrated complex. Kriegel contrasts the same-order view with the higher-order theory of consciousness, where a separate state represents the conscious mental state. He argues that the same-order view has distinct advantages, notably that it explains the intimate connection between conscious states and our awareness of them.

### *Chapter 14: "What Kind of Awareness Is Awareness of Awareness?" by Michelle Montague*

Montague explores and defends the thesis inspired by Franz Brentano that consciousness constitutively involves an awareness of itself. This special sort of awareness is part of the conscious state itself and is not a separate act of awareness. Montague notes that this "awareness of awareness" is peripheral, rather than the focus of our attention. Further, this awareness is nonconceptual—it does not involve bringing experience under the concept of awareness. Finally, it is a kind of representation not involving a separate representational target. It is essentially self-directed. Montague then contrasts this special sort of awareness with introspection, a process involving attentive conceptual resources directed at our conscious experience. She argues that the more basic form of awareness of awareness is distinct from introspection and cannot be transformed into introspection.

### *Chapter 15: "Higher-Order Theories of Consciousness," by Josh Weisberg*

Weisberg presents and defends the "higher-order theory" of mental state consciousness. On this view, a mental state is conscious when a subject represents themselves as being in it by way of a "higher-order" representation—a representation about other mental states. The basic motivation for the higher-order view is the "transitivity principle," the claim that mental states are conscious when we are appropriately conscious of them. Weisberg defends the status of the transitivity principle as the right way to fix the data that a theory of consciousness must explain. Then he presents the main versions of the higher-order theory: higher-order perception theory and higher-order thought theory. The chapter then considers some of the major objections to the higher-order theory, notably the worry that higher-order states might misrepresent us as being in states we are not actually in. The chapter closes by considering possible empirical evidence for the higher-order view, and by sketching how the theory deals with the explanatory gap and hard problem of consciousness.

## *Part V: Agency and Physicalism*

### *Chapter 16: "Perceptual Consciousness as Mental Activity," by Susanna Schellenberg*

Schellenberg develops and defends the claim that perceptual consciousness is constituted by mental activity. That is, we are perceptually conscious when we exercise the proper discriminatory capacity. This stands in contrast to views that hold that perceptual consciousness involves a special relation to either a perceived particular or to abstract objects like universals, properties, or intentional objects. A perceptual capacity is fixed by its function to successfully discriminate some object or property instantiation, but it can operate in the absence of such successful discrimination. Therefore, we get an explanation of hallucinatory or illusory experience that does not involve an instantiated relation to a particular. Schellenberg also shows how the capacity theory entails a moderate form of representationalism about consciousness while retaining its independence from the relational view.

### *Chapter 17: "The Proprietary Nature of Agentive Experience," by Myrto Mylopoulos*

Mylopoulos defends the claim that there is a distinctive form of experience present when we successfully engage in action. This propriety "agentive experience" is not reducible to other, more basic forms of experience, like the feeling of trying, of proprioception, or of desire. Mylopoulos argues for her claim by rebutting both strong and moderate skepticism of the view. The strong skeptic holds that there is no experiential feature at all

marking action. Instead, we only have experience of our actions being thwarted. That is, deliberately moving our hand does not feel like anything—we just do it. But if we try to move our hand and for some reason it fails to move, we notice that. The moderate skeptic holds that there is indeed a feeling accompanying successful action, but it is ultimately reducible to other features of experience and thus is not proprietary. Mylopoulos cites empirical cases where agents have deficits in their ability to perform actions to show that the proprietary view best captures the date to be explained.

### *Chapter 18: "Realistic Monism," by Galen Strawson*

Strawson argues that consciousness cannot be explained in terms of the combination of any nonexperiential physical elements. This poses the central dilemma of the mind-body problem: either physicalism must accept that there is no consciousness, because all there is is nonexperiential physical stuff, or we must accept dualism about experience. But Strawson contends that this dilemma rests on the claim that basic physical matter must be nonexperiential —that it cannot possess any experiential properties. This assumes an overly restrictive concept of the physical, one based on the mathematical approach of modern physics. But physics is silent about the intrinsic, nonmathematical features of basic physical matter. This leaves space to posit experiential properties at the intrinsic heart of the matter mathematically described by physics. Strawson then defends this "panpsychic" view, where all matter is partly experiential, against alternatives involving the "brute emergence" of consciousness at higher levels.

### *Chapter 19: "Property Dualism and the Merits of Solutions to the Mind-Body Problem," by Fiona Macpherson*

Macpherson critically evaluates Galen Strawson's arguments in his paper "Realistic Monism." She argues first that Strawson's characterization of his view as physicalist goes against common usage where the term explicitly rules out physical stuff as having irreducible experiential properties. She then argues that though Strawson's position can reasonably support a monism about substance—that there is just one kind of basic stuff that everything is composed of—it does not rule out property dualism and indeed is best interpreted this way. Property dualism holds that there are two kinds of fundamental property, experiential and physical, and neither can be explained in terms of the other. This view is generally opposed to physicalism and so Strawson's position is not best seen as a version of physicalism. She closes her paper by considering a number of worries for Strawson's panpsychism, including the problem of explaining how fundamental experiential elements might combine to form our subjective experiences.

### Reference

Galilei, Galileo (1623/1957) *The Assayer*, in *Discoveries and Opinions of Galileo*, trans. and ed. Stillman Drake, New York: Double Day and Co., pp. 237–238.

# PART I

# Problems of Consciousness

1

# What Is It Like to Be a Bat?

Thomas Nagel

Consciousness is what makes the mind-body problem really intractable. Perhaps that is why current discussions of the problem give it little attention or get it obviously wrong. The recent wave of reductionist euphoria has produced several analyses of mental phenomena and mental concepts designed to explain the possibility of some variety of materialism, psychophysical identification, or reduction.[1] But the problems dealt with are those common to this type of reduction and other types, and what makes the mind-body problem unique, and unlike the water-$H_2O$ problem or the Turing machine-IBM machine problem or the lightning-electrical discharge problem or the gene-DNA problem or the oak tree-hydrocarbon problem, is ignored.

Every reductionist has his favorite analogy from modern science. It is most unlikely that any of these unrelated examples of successful reduction will shed light on the relation of mind to brain. But philosophers share the general human weakness for explanations of what is incomprehensible in terms suited for what is familiar and well understood, though entirely different. This has led to the acceptance of implausible accounts of the mental largely because they would permit familiar kinds of reduction. I shall try to explain why the usual examples do not help us to understand the relation between mind and body—why, indeed, we have at present no conception of what an explanation of the physical nature of a mental phenomenon would be. Without consciousness the mind-body problem would be much less interesting. With consciousness it seems hopeless. The most important and characteristic feature of conscious mental phenomena is very poorly understood. Most reductionist theories do not even try to explain it. And careful examination will show that no currently

1 Examples are J. J. C. Smart, *Philosophy and Scientific Realism* (London, 1963); David K. Lewis, "An Argument for the Identity Theory," *Journal of Philosophy,* LXIII (1966), reprinted with addenda in David M. Rosenthal, *Materialism* & *the Mind-Body Problem* (Englewood Cliffs, N. J., 1971); Hilary Putnam, "Psychological Predicates" in Capitan and Merrill, *Art, Mind,* & *Religion* (Pittsburgh, 1967), reprinted in Rosenthal, *op. cit.*, as "The Nature of Mental States"; D. M. Armstrong, *A Materialist Theory of the Mind* (London, 1968); D. C. Dennett, *Content and Consciousness* (London, 1969). I have expressed earlier doubts in "Armstrong on the Mind," *Philosophical Review,* LXXIX (1970), 394–403; "Brain Bisection and the Unity of Consciousness," *Synthèse*, 22 (1971); and a review of Dennett, *Journal of Philosophy*, LXIX (1972). See also Saul Kripke, "Naming and Necessity" in Davidson and Harman, *Semantics of Natural Language* (Dordrecht, 1972), esp. pp. 334–342; and M. T. Thornton, "Ostensive Terms and Materialism," *The Monist*, 56 (1972).

Original publication details: Thomas Nagel, "What Is It Like to Be a Bat?" pp. 435–450 from *The Philosophical Review,* 83:4 (October 1974). Duke University Press, public domain.

*Consciousness*, First Edition. Edited by Josh Weisberg and David Rosenthal.

available concept of reduction is applicable to it. Perhaps a new theoretical form can be devised for the purpose, but such a solution, if it exists, lies in the distant intellectual future.

Conscious experience is a widespread phenomenon. It occurs at many levels of animal life, though we cannot be sure of its presence in the simpler organisms, and it is very difficult to say in general what provides evidence of it. (Some extremists have been prepared to deny it even of mammals other than man.) No doubt it occurs in countless forms totally unimaginable to us, on other planets in other solar systems throughout the universe. But no matter how the form may vary, the fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism. There may be further implications about the form of the experience; there may even (though I doubt it) be implications about the behavior of the organism. But fundamentally 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.

We may call this the subjective character of experience. It is not captured by any of the familiar, recently devised reductive analyses of the mental, for all of them are logically compatible with its absence. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved like people though they experienced nothing.[2] It is not analyzable in terms of the causal role of experiences in relation to typical human behavior—for similar reasons.[3] I do not deny that conscious mental states and events cause behavior, nor that they may be given functional characterizations. I deny only that this kind of thing exhausts their analysis. Any reductionist program has to be based on an analysis of what is to be reduced. If the analysis leaves something out, the problem will be falsely posed. It is useless to base the defense of materialism on any analysis of mental phenomena that fails to deal explicitly with their subjective character. For there is no reason to suppose that a reduction which seems plausible when no attempt is made to account for consciousness can be extended to include consciousness. Without some idea, therefore, of what the subjective character of experience is, we cannot know what is required of a physicalist theory.

While an account of the physical basis of mind must explain many things, this appears to be the most difficult. It is impossible to exclude the phenomenological features of experience from a reduction in the same way that one excludes the phenomenal features of an ordinary substance from a physical or chemical reduction of it—namely, by explaining them as effects on the minds of human observers.[4] If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view.

Let me first try to state the issue somewhat more fully than by referring to the relation between the subjective and the objective, or between the *pour-soi* and the *en-soi*. This is far from easy. Facts about what it is like to be an *X* are very peculiar, so peculiar that some may be inclined to doubt their reality, or the significance of claims about them. To illustrate the connection between subjectivity and a point of view, and to make evident the importance of subjective features, it will help to explore the matter in relation to an example that brings out clearly the divergence between the two types of conception, subjective and objective.

I assume we all believe that bats have experience. After all, they are mammals, and there is no more doubt that they have experience than that mice or pigeons or whales have experience. I have chosen bats instead of wasps or flounders because if one travels too far down the phylogenetic tree, people gradually shed their faith that there is experience there at all. Bats, although more closely related to us than those other species, nevertheless present a range of activity and a sensory apparatus so different from ours that the problem I want to pose is exceptionally vivid (though it certainly could be raised with other species). Even without the benefit of philosophical reflection, anyone who has spent some time in an enclosed space with an excited bat knows what it is to encounter a fundamentally *alien* form of life.

2 Perhaps there could not actually be such robots. Perhaps anything complex enough to behave like a person would have experiences. But that, if true, is a fact which cannot be discovered merely by analyzing the concept of experience.

3 It is not equivalent to that about which we are incorrigible, both because we are not incorrigible about experience and because experience is present in animals lacking language and thought, who have no beliefs at all about their experiences.

4 Cf. Richard Rorty, "Mind-Body Identity, Privacy, and Categories," *The Review of Metaphysics*, XIX (1965), esp. 37–38.

I have said that the essence of the belief that bats have experience is that there is something that it is like to be a bat. Now we know that most bats (the microchiroptera, to be precise) perceive the external world primarily by sonar, or echolocation, detecting the reflections, from objects within range, of their own rapid, subtly modulated, high-frequency shrieks. Their brains are designed to correlate the outgoing impulses with the subsequent echoes, and the information thus acquired enables bats to make precise discriminations of distance, size, shape, motion, and texture comparable to those we make by vision. But bat sonar, though clearly a form of perception, is not similar in its operation to any sense that we possess, and there is no reason to suppose that it is subjectively like anything we can experience or imagine. This appears to create difficulties for the notion of what it is like to be a bat. We must consider whether any method will permit us to extrapolate to the inner life of the bat from our own case,[5] and if not, what alternative methods there may be for understanding the notion.

Our own experience provides the basic material for our imagination, whose range is therefore limited. It will not help to try to imagine that one has webbing on one's arms, which enables one to fly around at dusk and dawn catching insects in one's mouth; that one has very poor vision, and perceives the surrounding world by a system of reflected high-frequency sound signals; and that one spends the day hanging upside down by one's feet in an attic. In so far as I can imagine this (which is not very far), it tells me only what it would be like for *me* to behave as a bat behaves. But that is not the question. I want to know what it is like for a *bat* to be a bat. Yet if I try to imagine this, I am restricted to the resources of my own mind, and those resources are inadequate to the task. I cannot perform it either by imagining additions to my present experience, or by imagining segments gradually subtracted from it, or by imagining some combination of additions, subtractions, and modifications.

To the extent that I could look and behave like a wasp or a bat without changing my fundamental structure, my experiences would not be anything like the experiences of those animals. On the other hand, it is doubtful that any meaning can be attached to the supposition that I should possess the internal neurophysiological constitution of a bat. Even if I could by gradual degrees be transformed into a bat, nothing in my present constitution enables me to imagine what the experiences of such a future stage of myself thus metamorphosed would be like. The best evidence would come from the experiences of bats, if we only knew what they were like.

So if extrapolation from our own case is involved in the idea of what it is like to be a bat, the extrapolation must be incompletable. We cannot form more than a schematic conception of what it *is* like. For example, we may ascribe general *types* of experience on the basis of the animal's structure and behavior. Thus we describe bat sonar as a form of three-dimensional forward perception; we believe that bats feel some versions of pain, fear, hunger, and lust, and that they have other, more familiar types of perception besides sonar. But we believe that these experiences also have in each case a specific subjective character, which it is beyond our ability to conceive. And if there is conscious life elsewhere in the universe, it is likely that some of it will not be describable even in the most general experiential terms available to us.[6] (The problem is not confined to exotic cases, however, for it exists between one person and another. The subjective character of the experience of a person deaf and blind from birth is not accessible to me, for example, nor presumably is mine to him. This does not prevent us each from believing that the other's experience has such a subjective character.)

If anyone is inclined to deny that we can believe in the existence of facts like this whose exact nature we cannot possibly conceive, he should reflect that in contemplating the bats we are in much the same position that intelligent bats or Martians[7] would occupy if they tried to form a conception of what it was like to be us. The structure of their own minds might make it impossible for them to succeed, but we know they would be wrong to conclude that there is not anything precise that it is like to be us: that only certain general types of mental state could be ascribed to us (perhaps perception and appetite would be concepts common to us both; perhaps not). We know they would be wrong to draw such a skeptical conclusion because we know what it is like to be us.

5 By "our own case" I do not mean just "my own case," but rather the mentalistic ideas that we apply unproblematically to ourselves and other human beings.

6 Therefore the analogical form of the English expression "what it is *like*" is misleading. It does not mean "what (in our experience) it *resembles*," but rather "how it is for the subject himself."

7 Any intelligent extraterrestrial beings totally different from us.

And we know that while it includes an enormous amount of variation and complexity, and while we do not possess the vocabulary to describe it adequately, its subjective character is highly specific, and in some respects describable in terms that can be understood only by creatures like us. The fact that we cannot expect ever to accommodate in our language a detailed description of Martian or bat phenomenology should not lead us to dismiss as meaningless the claim that bats and Martians have experiences fully comparable in richness of detail to our own. It would be fine if someone were to develop concepts and a theory that enabled us to think about those things; but such an understanding may be permanently denied to us by the limits of our nature. And to deny the reality or logical significance of what we can never describe or understand is the crudest form of cognitive dissonance.

This brings us to the edge of a topic that requires much more discussion than I can give it here: namely, the relation between facts on the one hand and conceptual schemes or systems of representation on the other. My realism about the subjective domain in all its forms implies a belief in the existence of facts beyond the reach of human concepts. Certainly it is possible for a human being to believe that there are facts which humans never *will* possess the requisite concepts to represent or comprehend. Indeed, it would be foolish to doubt this, given the finiteness of humanity's expectations. After all, there would have been transfinite numbers even if everyone had been wiped out by the Black Death before Cantor discovered them. But one might also believe that there are facts which *could* not ever be represented or comprehended by human beings, even if the species lasted forever—simply because our structure does not permit us to operate with concepts of the requisite type. This impossibility might even be observed by other beings, but it is not clear that the existence of such beings, or the possibility of their existence, is a precondition of the significance of the hypothesis that there are humanly inaccessible facts. (After all, the nature of beings with access to humanly inaccessible facts is presumably itself a humanly inaccessible fact.) Reflection on what it is like to be a bat seems to lead us, therefore, to the conclusion that there are facts that do not consist in the truth of propositions expressible in a human language. We can be compelled to recognize the existence of such facts without being able to state or comprehend them.

I shall not pursue this subject, however. Its bearing on the topic before us (namely, the mind-body problem) is that it enables us to make a general observation about the subjective character of experience. Whatever may be the status of facts about what it is like to be a human being, or a bat, or a Martian, these appear to be facts that embody a particular point of view.

I am not adverting here to the alleged privacy of experience to its possessor. The point of view in question is not one accessible only to a single individual. Rather it is a *type*. It is often possible to take up a point of view other than one's own, so the comprehension of such facts is not limited to one's own case. There is a sense in which phenomenological facts are perfectly objective: one person can know or say of another what the quality of the other's experience is. They are subjective, however, in the sense that even this objective ascription of experience is possible only for someone sufficiently similar to the object of ascription to be able to adopt his point of view—to understand the ascription in the first person as well as in the third, so to speak. The more different from oneself the other experiencer is, the less success one can expect with this enterprise. In our own case we occupy the relevant point of view, but we will have as much difficulty understanding our own experience properly if we approach it from another point of view as we would if we tried to understand the experience of another species without taking up *its* point of view.[8]

This bears directly on the mind-body problem. For if the facts of experience—facts about what it is like *for* the experiencing organism—are accessible only from one point of view, then it is a mystery how the true character of experiences could be revealed in the physical operation of that organism. The latter is a domain of objective

8 It may be easier than I suppose to transcend inter-species barriers with the aid of the imagination. For example, blind people are able to detect objects near them by a form of sonar, using vocal clicks or taps of a cane. Perhaps if one knew what that was like, one could by extension imagine roughly what it was like to possess the much more refined sonar of a bat. The distance between oneself and other persons and other species can fall anywhere on a continuum. Even for other persons the understanding of what it is like to be them is only partial, and when one moves to species very different from oneself, a lesser degree of partial understanding may still be available. The imagination is remarkably flexible. My point, however, is not that we cannot *know* what it is like to be a bat. I am not raising that epistemological problem. My point is rather that even to form a *conception* of what it is like to be a bat (and a fortiori to know what it is like to be a bat) one must take up the bat's point of view. If one can take it up roughly, or partially, then one's conception will also be rough or partial. Or so it seems in our present state of understanding.