

O W E N F L A N A G A N

Consciousness

Consciousness

Reconsidered

Consciousness Reconsidered

Consciousness Reconsidered

Owen Flanagan

**A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England**

© 1992 Massachusetts Institute of Technology

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

This book was set in Palatino by Asco Trade Typesetting Ltd., Hong Kong, and was printed and bound in the United States of America.

Fourth printing, 1998

Library of Congress Cataloging-in-Publication Data
Flanagan, Owen J.

Consciousness reconsidered / Owen Flanagan.

p. cm.

Includes bibliographical references.

ISBN 0-262-06148-1 (hb), 0-262-56077-1 (pb)

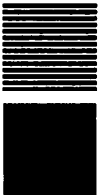
1. Consciousness. 2. Mind and body. I. Title.

B808.9.F57 1992

126—dc20

92-10057

CIP



MIT Press

0262560771

Flanagan

Consciousness Reconsidered



*To my sisters—Virginia, Kathleen, and Nancy—my brother Mark, and the
memory of our brother Peter*

Contents

Preface xi

Chapter 1

Subjectivity and the Natural Order	1
1 <i>Subjectivity</i>	1
2 <i>Philosophical Space</i>	1
3 <i>Consciousness and Cognitivism</i>	3
4 <i>Conscious Shyness and the New Mysterianism</i>	8
5 <i>Subjectivity, Objectivity, and the Natural Method</i>	11
6 <i>The Natural Method: Three Examples</i>	13

Chapter 2

Quining Consciousness	21
1 <i>What Quining Is</i>	21
2 <i>The Philogiston Objection</i>	22
3 <i>The Karma Objection</i>	28

Chapter 3

Consciousness and the Brain	35
1 <i>Mind, Brain, and Experience</i>	35
2 <i>The Evolution of Darwin Machines</i>	40
3 <i>Neural Darwinism</i>	46
4 <i>Self and Nonself</i>	49
5 <i>Sensory Qualia and Neural Vectors</i>	51
6 <i>Experiential Sensitivity versus Informational Sensitivity</i>	55
7 <i>Identity Theory to a First Approximation?</i>	56

Chapter 4

Qualia	61
1 <i>Quining Qualia</i>	61
2 <i>What Qualia Are</i>	62
3 <i>Sensational and Nonsensational Qualia</i>	67
4 <i>Inverted Qualia</i>	68
5 <i>Intractable Qualia</i>	72

Chapter 5

The Missing Shade of You	87
1 <i>The Subject Is Experience</i>	87
2 <i>The Missing Shade of Blue</i>	87
3 <i>The Missing Shade of You</i>	89
4 <i>Capturing You: The Very Idea</i>	89
5 <i>Grasping Experiences</i>	94
6 <i>Missing Shades Again</i>	97
7 <i>Refrain: The Missing Shade of You</i>	102

Chapter 6

The Mystery of Consciousness	109
1 <i>The Mystery of Qualia</i>	109
2 <i>The First Argument: Why We Can't Solve the Problem of Consciousness</i>	111
3 <i>Critique of the First Argument</i>	112
4 <i>Some Further Objections and Replies</i>	118
5 <i>The Second Argument: The Hidden Structure of Consciousness</i>	120
6 <i>Critique of the Hidden-Structure Argument</i>	123

Chapter 7

Conscious Inessentialism and the Epiphenomenalist Suspicion	129
1 <i>Conscious Inessentialism</i>	129
2 <i>The Epiphenomenalist Suspicion</i>	130
3 <i>Some Arguments from Design</i>	132
4 <i>An Experiment in Epiphenomenalism</i>	136
5 <i>Teleological Functionalism, Epiphenomenalism, and Defects of Consciousness</i>	139
6 <i>Blocking Teleological Functionalism</i>	145
7 <i>Phenomenal Access</i>	147
8 <i>Can the Epiphenomenalist Explain Everything That the Teleological Functionalist Can Explain?</i>	149

Chapter 8

The Stream of Consciousness	153
1 <i>What Is It Like to Have a Normal Consciousness?</i>	153
2 <i>Phenomenology</i>	153
3 <i>Are We Ever Wholly Unconscious?</i>	156
4 <i>The Stream</i>	158
5 <i>The Function of the Stream</i>	166
6 <i>Nonstandard Phenomenologies</i>	166
7 <i>Phenomenal Competition</i>	169

Chapter 9

The Illusion of the Mind's "I"	177
1 <i>The Mind's "I"</i>	177
2 <i>The Illusion of Ego</i>	177
3 <i>Me: The Self as Known</i>	179
4 <i>The Seduction</i>	182
5 <i>I: The Self as Knower</i>	183
6 <i>Personal Identity</i>	187
7 <i>The Unbearable Lightness of Egolessness</i>	181

Chapter 10

Consciousness and the Self	193
1 <i>Weak and Strong Self-Consciousness</i>	193
2 <i>Actual and Self-Represented Identity</i>	195
3 <i>The Narrative Structure of Self-Representation</i>	197
4 <i>Self-Emergence</i>	200
5 <i>Whom the Self Plays For</i>	202
6 <i>The Self as Fiction</i>	204
7 <i>Real and Unreal Selves</i>	207

Chapter 11

A Unified Theory of Consciousness?	213
1 <i>Prospects for a Theory of Consciousness</i>	213
2 <i>A Brief Recapitulation</i>	220

References 223

Index 229

Preface

Naturalism is the view that the mind-brain relation is a natural one. Mental processes just are brain processes. This idea has increased in credibility as the picture of the mind as a sophisticated information processor has become more widely accepted and as neuroscience has shown that the brain possesses both the complexity and the power to do the information processing human minds in fact do.

But there is a gnawing suspicion that the picture of persons as sophisticated information processors leaves something out. And indeed, it does. We are *conscious creatures*. Perhaps we are information processors, but if we are, we are *conscious* information processors. Our mental life has a phenomenal side, a subjective side, that the most sophisticated information processor might lack.

Whereas the brain seems suited to processing information, it is harder to imagine the brain's giving rise to consciousness. The very idea of consciousness materializing, of subjectivity being realized in the activity of a physical organism, is puzzling. The rich phenomenology of the conscious stream and complex neural activity appear to belong to two entirely different orders: the subjective and the objective. This book is an attempt to make less puzzling the idea that consciousness is a natural phenomenon. I present a view of consciousness that I call "constructive naturalism." Consciousness is a natural phenomenon, and we can construct a theory about its nature, forms, roles, and origins by blending insights from phenomenology, psychology, cognitive science, neuroscience, and evolutionary biology. Consciousness is neither miraculous nor terminally mysterious.

The origins of this book date back to the seminar "Consciousness" that I taught in the spring of 1989. Thanks to a grant from the Alfred P. Sloan Foundation, I was able to invite as visitors Daniel Dennett, Michael Gazzaniga, Patricia Kitcher, Carolyn Ristau, Georges Rey, David Rosenthal, and Robert Van Gulick. Two colleagues from Wellesley's psychology department and fellow members of our cognitive science group, Margery Lucas and Larry Rosenblum (now at the University of California, Riverside), and my colleague from philosophy, Ken Winkler, also participated in the seminar. The seminar was a source of great intellectual stimulation to me,

thanks to these individuals and to the wonderful group of students who were involved in the exploration.

In the summer of 1989, as I was gathering my thoughts from the seminar, Colin McGinn's important paper "Can We Solve the Mind-Body Problem?" appeared in *Mind*. McGinn argued that the consciousness-brain problem could never be solved by creatures with our kind of minds. McGinn's argument was imaginative, forceful, and complex—as I discovered when he was kind enough to send me several chapters of his then forthcoming book, *The Problem of Consciousness* (1991). If McGinn was right, then my whole approach to the problem of consciousness was just plain wrong. So McGinn's work inspired me. I first presented a response to his argument to my department's faculty seminar in the fall of 1989. Then during the winter and spring of 1990, I sketched out my view on the problem of consciousness and offered a reply to McGinn's 1989 paper. I published this material as the chapter "Consciousness" in the second edition of *The Science of the Mind* (1991). Writing that chapter was a good first start, but when I finished it, I found that I had a good deal more to say about consciousness. *Consciousness Reconsidered* was conceived at that time. Part of the unfinished work involved providing a more complete response to McGinn's arguments against the possibility of constructive naturalism. Chapter 6 below is specifically devoted to this task. But the book as a whole is the real response to McGinn's anticonstructive naturalism. For its overall goal is to show that a constructive theory of consciousness is not merely possible; in certain respects it is already actual.

During the fall term of the 1990/1991 academic year, Daniel Dennett invited me to participate in his seminar "Consciousness" at Tufts. Weekly conversations with Dan about his almost completed book *Consciousness Explained* (1991a) and with Ned Block and Andrew Woodfield were especially helpful in sharpening my views and giving me the courage of my convictions.

I started to write again during the spring term of 1991, when I taught a new version of my "Consciousness" seminar. This time no foundation paid for visitors, so I forged ahead with a new group of students, as intelligent and conscientious as the first group had been. I thank them for helping me get clearer on many of the issues discussed in these pages. My students and Hu-Min Ji, a visiting philosopher from Beijing, forced me to explain how, where, and why my views differ from those of other naturalists like Daniel Dennett, Robert Van Gulick, Paul Churchland, Patricia Churchland, and Colin McGinn. And they pressed me hard to respond to challenges to naturalism itself, in particular, to the challenges of Thomas Nagel and Frank Jackson.

I am indebted to a great many people. Over the past few years, I have had the good fortune to talk extensively about the problem of conscious-

ness with Robert Van Gulick. Bob's excellent work on consciousness increased my confidence in the idea that insights from the phenomenological, psychological, and neuroscientific levels can be blended profitably in seeking to understand the nature and function of consciousness. In addition to Bob Van Gulick, I also had help from Ned Block, Robert McCauley, Daniel Dennett, George Graham, Paul Churchland, Patricia Churchland, and Jonathan Cheek, who gave me encouragement, good leads, helpful preprints and reprints, and sage advice. A visit to Saint Louis came at just the right time. There Roger Gibson introduced me to Marc Raichle, from whom I received an immensely helpful tutorial on the latest in brain-imaging research over the course of a most pleasant dinner.

David Galloway read the first draft and made many helpful comments. Ned Block made me think extra hard about epiphenomenalism and the problem of the function of consciousness by letting me read some of his unpublished work on the topic and with his written comments on chapter 7. George Graham read and commented on the entire manuscript. George pressed me to provide a clear map of the various philosophical positions on the problem of consciousness. George described his own impression of the live positions in a most helpful manner. His map helped me immensely in drawing my own map of "philosophical space" in the second section of chapter 1. That section is dedicated to George. Ken Winkler and Jerry Samet read the manuscript, sometimes successive versions, with the utmost care. Both wrote me pages and pages of comments and spent many hours talking the project over with me. I feel most fortunate to have two such excellent and tireless critics as friends. I am eternally grateful to both of them for helping me make this book far better than it would otherwise have been.

Joyce Knowlton Walworth and our children, Ben and Kate, helped make writing this book an especially pleasant experience. Writing is an activity I do at home in the presence of my family. Ben and Kate always ask me what I am working on, and as things that I work on go, they found this topic unusually fun and interesting. Ben specializes in the area of consciousness and sports. Apparently, when a Little League pitcher stares at his catcher, there is heightened awareness as the next pitch is planned. Once the plan is firmed up, automaticity takes hold, and a split-finger fastball or change-up is thrown. If consciousness kicks in at the wrong point, it is called a "wild pitch." Things are somewhat different for a second baseman or shortstop. An infielder has clear awareness of the sound of the ball coming off the bat; there is a vivid visual lock onto the ball as it heads in his direction, but there is only dim awareness of what his body is doing as it moves to make the catch. If body awareness is heightened, it is called an "error." Soccer and basketball have their own distinctive phenomenology. Whereas Ben works primarily in phenomenology and on the connection

between types of consciousness and behavior, Kate is a neurophilosopher, with expertise on the sensory nerves. Consciousness is in the brain. You feel the pain in your toe not because the pain is in the toe but because the nerves in the toe tell the brain to tell you that something is wrong in your toe. They don't teach this in second grade. But Kate knows it, and she hasn't even read *Neurophilosophy* or *Consciousness Explained*.

This book is dedicated to my brothers and sisters. They each in their own way command my deep love, respect, and admiration. This book is a small token of appreciation for all they mean to me individually and collectively.

Consciousness Reconsidered

Chapter 1

Subjectivity and the Natural Order

1 *Subjectivity*

Three of the greatest perplexities are these. Why is there something rather than nothing? How did some of the stuff there come to be alive? How did some of the living stuff come to be conscious? Alongside and intimately related to the questions of how and why matter, life, and consciousness came into being are questions about the nature of matter, life, and consciousness.

Here I take on the third perplexity and sketch a naturalistic theory of consciousness. My aim is to say something illuminating about the nature, function, and origin of consciousness. Subjectivity has emerged so far only in certain biological systems. It makes sense, therefore, to seek a theory of consciousness with the guidance of the neo-Darwinian theory of evolution and the best current brain science.

2 *Philosophical Space*

There are several main philosophical positions on the problem of consciousness. First, there is *nonnaturalism*, the view that consciousness is not a natural phenomenon and therefore cannot be understood in naturalistic terms. Some nonnaturalists think that consciousness can be made intelligible if it is understood as a power of a nonphysical substance or as composed of nonphysical properties (Popper and Eccles 1977). Others think that we need to invoke a supernatural cause to explain why phenomenal qualia, the sensation of red or the scent of a rose, are correlated with specific types of brain states (Adams 1987, Swinburne 1984). Still others think that consciousness is miraculous. Like transubstantiation and the Trinity, it is not for us to fathom.

Second, there is *principled agnosticism* (Nagel 1974, 1986). Naturalism is a position we do not understand, because we do not understand (at least at present) how the relation of consciousness and the brain can be made intelligible in naturalistic terms. We don't understand what it would mean to give an objective account of subjectivity. Since one should not

believe a theory one does not even understand, agnosticism is the best policy.

Third, there is *anticonstructive naturalism*, noumenal naturalism, or the *new mysterianism*, as I will also call it (McGinn 1991). This is the view that naturalism is true. There are in fact properties of the brain that account naturalistically for consciousness. But we cannot grasp these properties or explain how consciousness depends on them. Consciousness is terminally mysterious to our minds but possibly not to minds of greater intelligence. It is terminally mysterious not because it is a nonnatural phenomenon, not because it is a miracle, but because an understanding of its nature is "cognitively closed" to us. The problem of consciousness is a case where we know how to ask the question but lack the mental powers to find the answer.

Fourth, there is *eliminativist naturalism* (P. M. Churchland 1981, P. S. Churchland 1983). According to the eliminativist, naturalism is true. The complete story of our brain will tell the complete story of our mental life. But there is a sense in which consciousness cannot be explained. Consciousness is a concept that is simultaneously too simplistic, too vague, and too historically embedded in false and confused theory to perspicuously denote a phenomenon or set of phenomena in need of explanation. Concepts like consciousness, qualia, and subjectivity are unhelpful in setting out the explanatory agenda for a naturalistic theory of mind. Whatever genuine phenomena these concepts inchoately gesture toward will be explained by the science of the mind. But the explanation will proceed best if we eliminate these concepts from the explanatory platter and seek more perspicuous and credible replacements undergirded by a rich neuroscientific theory.

Finally, there is *constructive naturalism*. This is the position I aim to defend. Like the anticonstructivist and the eliminativist, I think that naturalism is true. Against the anticonstructivist and principled agnostic, I maintain that there is reason for optimism about our ability to understand the relation between consciousness and the brain. We can make intelligible the existence of consciousness in the natural world. Against the eliminativist, I maintain that the concept of consciousness, despite its shortcomings, is needed, at least at the beginning of inquiry, to mark what is in need of explanation. Phenomenal, qualitative consciousness is what needs to be explained. Indeed, the method I propose for attacking the problem of consciousness requires that we use, but treat as revisable, our ordinary first-person modes for taxonomizing subjectivity.

Even at this early stage in the development of the science of the mind, there are deep differences of opinion among naturalists about whether the mystery of consciousness can be made to yield, about whether there are such things as phenomenal consciousness and qualia in need of explanation,

about the importance of consciousness in the overall economy of mind, and about what shape the theory will take and what methods will be used to construct it. This essay is an attempt to contribute to the debate among naturalists. My aim is to show that constructive naturalism is a better view than anticonstructive naturalism, eliminative naturalism, or principled agnosticism.

Happily, I am not alone in believing that a constructive theory is possible. Recent work by P. S. Churchland (1986), P. M. Churchland (1989), and Daniel Dennett (1991a) is in the mode of constructivist naturalism. All three take conscious experience seriously as a phenomenon or set of phenomena to be explained. No one now defends the outright elimination of our commonsense ways of conceiving of mind. To be sure, the Churchlands doubt that folk psychology will fare well as inquiry proceeds, and they expect the real illumination to be provided by neuroscience rather than by cognitive psychology or philosophy. And Dennett has his eliminativist moments. For example, he favors elimination of the concept of qualia on the grounds that it names nothing (1988c). But Dennett is not as charmed as most traditional eliminativists by the promise of neuroscience. He instead favors different styles of explanation for different phenomena and proposes a "mild realism" about the patterns our commonsense psychological vocabulary picks out (1991d). The disagreements within constructive naturalism are plentiful. The important point is that these disagreements proceed in a context of agreement that mind in general and consciousness in particular will yield their secrets only by coordinating all our informational sources at once.

I would like to think that this essay might move nonnaturalists. Perhaps it will make them see that naturalism is more resourceful than they thought. But it is important to acknowledge at the start that I provide no argument against nonnaturalism. I do not think that nonnaturalism is an incoherent view, and I am familiar enough with the dialectical moves of nonnaturalists to know that from their point of view no naturalistic story will satisfy. I will be satisfied if within naturalism the constructivist position gains credibility.

3 *Consciousness and Cognitivism*

Behaviorism, positivism's close kin and psychology's prude, fell from power during the 1960s, and psychologists renewed the study of mind with great energy and freedom. The new regime was cognitivism. The irony is that the return of mind to psychology attending the demise of behaviorism and the rise of cognitivism did not mark the return of consciousness to the science of the mind. Mind without consciousness? How is that possible?

In the first place, the rejection of behaviorism did not take place with complete methodological abandon. A certain appropriate positivistic re-

serve remained. It was one thing to draw inferences about the decay rates of visual images after brief exposures and about the structure of short-term memory and its relations to long-term memory by way of sensitive tachistoscopic experiments, reaction-time tests, and retention experiments. It was a wholly other thing to try to make scientific pronouncements about consciousness. Consciousness, after all, had press of this sort:

O, what a world of unseen visions and heard silences, this insubstantial country of mind! What ineffable essences, these touchless rememberings and unshowable reveries! And the privacy of it all! A speechless theatre of speechless monologue and prevenient counsel, an invisible mansion of moods, musings, and mysteries, an infinite resort of disappointments and discoveries. A whole kingdom where each of us reigns reclusively alone, questioning what we will, commanding what we can. A hidden hermitage where we may study out the troubled book of what we have done and yet may do. An introcosm that is more myself than anything I can find in the mirror. This consciousness that is myself of selves, that is everything, and yet nothing at all. (Jaynes 1976, 1)

Now this seemed like a good thing for a psychologist interested in gaining a reputation *not* to study. Whereof one cannot speak, one should be silent. What is essentially private is something about which science should offer no opinion. "Invisible mansions" of moods, musings, and mysteries, as appealing as they might be to mushy-minded mystics, hold no interest to the scientist of the mind.

In the second place, it seemed that one could map the mind, could provide a theory of intelligent mental life without committing oneself to any general view about the nature, function, or role of consciousness. To be sure, much work assumed that a certain amount of conscious mental activity was taking place in the domain under study. For example, research on imagistic rotation accepted that most people in fact experience themselves rotating picturelike mental objects. But the study of imagistic rotation was designed in part to test the accuracy of these conscious experiential reports. These experiences were not considered closed off to verbalization, nor did the verbalizations stand as the last word on underlying processes. The alleged "ineffable essences" were not ineffable. Despite the lack of a positive theory of the nature of consciousness, the study of specific cognitive domains that typically involve conscious experiences in those domains proceeded in a spirit of skepticism about the whole idea of inner theaters, introcosms, and realms that are in principle inaccessible.

For many, the hope was that the rhapsodic received view, with its ghostly resonances, would eventually come undone. A picture of conscious mental life would eventually emerge, but it would emerge in a piecemeal

fashion, associated with deepened understanding of specific cognitive domains, and alongside rejection of the idea that there exists a specific faculty of consciousness. This would take time, however, and in the meantime no respectable, or at least no respected, psychologist was trying to provide a general theory of consciousness. The general attitude during the rebirth of cognitive psychology is aptly described by Daniel Dennett: "Consciousness appears to be the last bastion of occult properties, epiphenomena, immeasurable subjective states—in short, the one area of mind best left to the philosophers who are welcome to it. Let them make fools of themselves trying to corral the 'quicksilver of phenomenology' into respectable theory" (1978, 149). A few years later, Dennett wrote, "Consciousness is making a comeback in psychology, but there is still residual skepticism, anxiety, and confusion about how to approach this perilous phenomenon scientifically" (1982, 159).

There was a third and in some ways philosophically more interesting reason of why the renewed excitement about the science of the mind was not accompanied by the reemergence of a theory of consciousness. It was widely noticed in many domains that the project of mapping out the complex information flows and networks constituting a mind, be that done at the brain level or the level of cognition, could be done without bringing consciousness into the story. Language acquisition and semantic processing, face recognition, and the like, seemed to proceed largely unconsciously. Furthermore, the dominant philosophical theory of mind, *computational functionalism*, had powerful affinities with the thesis of conscious inessentialism. *Conscious inessentialism* is the view that for any intelligent activity i , performed in any cognitive domain d , even if we do i with conscious accompaniments, i can in principle be done without these conscious accompaniments.

Most skeptics of strong artificial intelligence, the view that computers can in principle display any mental ability we display, press worries that machines cannot be given consciousness. Conscious inessentialism can be read as making this objection irrelevant. Mind does not require consciousness. All the input-output relations between stimuli and responses could be preserved in a system that passed the toughest conceivable Turing test, a test in which a computer must produce behavior indistinguishable from that of a person. To be sure, such a system might totally lack experience. But that simply proves the point: consciousness is not essential to intelligent mentality.

One can see how commitment to conscious inessentialism makes respectable the idea of a science of the mind that pays no significant attention to consciousness. It also explains why AI workers, be they members of aboriginal von Neumann tribes, true believers in the power of serial computation, or young connectionist upstarts devoted to massive parallelism,

are not obsessed with finding the right wiring diagrams to make consciousness emerge. Most everyone has abandoned the old illusion that consciousness is definitive of the mental, that each and every mental event is a conscious one. AI workers also think that a certain kind of species chauvinism has made us hang onto a second illusion. We think that consciousness, even if not involved in all mental activity, is essentially involved in our intelligent and purposeful activity. But it is not, since, the argument goes, any set of input-output relations can be subserved by an endless multiplicity of internal connections, and thus the toughest conceivable Turing test can be passed by systems totally lacking consciousness.

I reject conscious inessentialism. Consciousness *is* essentially involved in being intelligent and purposeful in the way(s) in which we are. Computational functionalism, in part because it normally involves commitment to conscious inessentialism, is the wrong sort of functionalism for the philosopher of psychology to be committed to. It would not be the wrong sort of view if the question were, What sorts of systems can preserve the input-output relations that we take to be (partly definitive) of intelligence and purposefulness? It would not be the wrong answer to that question because the answer to that question is simply the long disjunctive list of possible mechanical and biological models that could realize the proper input-output relations. Rather, computational functionalism is the wrong sort of view for the philosopher of mind to be committed to if he or she sees the project as continuous with that of explaining how intelligent, purposeful life occurs in humans and other earthly creatures. So long as there are tastes and smells and thoughts, there are conscious events to be explained. If these conscious events are not epiphenomenal, then they need to be mentioned in our full explanations. Furthermore, it is crucial to emphasize that neither input-output equivalence nor even some richer form of computational equivalence is the same as strict mental equivalence. Our behavioral outputs are actions; that is, they are bodily movements identified and individuated in part by the intentions and motives that constitute them. When we drive, we *signal* the intention to turn. When we laugh at a joke, we are *enjoying* the joke. Our robotic equivalent that drives just like us but lacks a motivational economy merely moves the turn signal. Our robotic equivalent that laughs at a joke but lacks a complex cognitive and conative economy does not enjoy it or think it funny.

I take the nature and function of conscious mental events and the actions they figure in as fundamental. For this reason I favor a form of functionalism that analyzes input-output relations in terms of the processes that mediate and subserve them in the normal biological cases, not in any possible cases whatsoever (Van Gulick 1988, 1990).

There is a fourth reason, partly related to and partly independent of the previous reasons, why the reemergence of mind did not result in conscious-

ness returning to its former pride of place. Call it the *epiphenomenalist suspicion*. In many domains in which there are in fact conscious accompaniments of the mental activity under study, it was noticed that those conscious accompaniments did not play any obvious or important causal role. First-person, phenomenological reports of the way things seem provided interesting data about experiences of imagery, short-term memory in matching tests, and so on. But the reports of conscious experience often seemed, in Dennett's wonderful analogy, to be like the press releases of the government spokesman who is out of the decision-making loop himself. A government spokesman is often one of the last persons inside the government to know what is going on. His role is quite unlike that of a truly important government functionary, the Secretary of Defense or State, for example, who exerts a firm hand in what is going on at each crucial choice point. Indeed, the press agent may, unbeknownst to himself, be engaged in massive dissembling—strategically misrepresenting the facts at the behest of those in the know. Other times he may mispeak, not so much because he is being called upon to play the role of mispeaker, but because he is too many communicative links away from the original action, like the fourth or fifth player in a game of “telephone.”

The emergence of cognitive science has been accompanied by the surprising reappearance of the sort of epiphenomenalism found attractive a century before, during the birth of the science of the mind. Recall Thomas Huxley's brazen pronouncement that the “soul stands to the body as the bell of a clock to the works, and consciousness answers to the sound which the bell gives out when it is struck. . . . We are conscious automata” (quoted in James 1890, 131). Perhaps consciousness stands to the mind-body as the President's press agent stands to the workings of the Oval Office, the Departments of State and Defense, the National Security Council, and the Office of Management and Budget. The fact that the press agent's speech acts occur *after* the important causal activity and decision making have taken place insinuates the possibility that consciousness may play an analogous role in the governance of our individual lives.

One might resist the analogy by insisting that consciousness simply is not like the President's press agent. It is like the President himself. But this is precisely the unquestioned assumption being challenged. Furthermore, even if we accept that consciousness has a presidential role, it is important to remember that Presidents differ greatly in the amount of control they exert. Many think that Ronald Reagan had, by historical standards, relatively little control of the government. According to this view, Reagan was the entertaining and eloquent spokesperson for a cadre of smart and hard-working powers (actually layers of powers), some known to outsiders, some unknown. This is not to deny that Reagan felt as if he were in charge in his role as “The Great Communicator,” first as Governor of California

and eventually as President. The point is that one can feel presidential, and indeed even *be* the President, but still be less in control than it seems from either the inside or the outside.

The epiphenomenalist suspicion is healthy up to a point. This is because we tend both to think that more of mental life is conscious than it in fact is, and because we tend to overstate the causal role of consciousness. But epiphenomenalism is ultimately implausible. There are, we will see, various good reasons to ascribe consciousness a function.

4 *Conscious Shyness and the New Mysterianism*

So far I have isolated four reasons why the reemergence of the study of mind associated with the rise of cognitivism did not result in the simultaneous reemergence of grand theorizing about consciousness. Each reason gives rise to a distinctive form of "conscious shyness." First, there was a certain methodological scrupulousness, a vestige of behaviorism's draconian restrictions on the study of the "inner" and the "unobservable." Second, there was the view that one could analyze various aspects of mentality with a background commitment to the existence of conscious experience in the domain under study but without a commitment to any general theory of the nature and function of consciousness. Third, there was the belief, inspired by computational functionalism, that consciousness was not in fact metaphysically essential. The thesis of conscious inessentialism says that there are possible worlds in which creatures computationally identical to us do not possess consciousness. Fourth, there is the epiphenomenalist suspicion, the suspicion that the role of consciousness may be far more limited in our case than consciousness itself (not surprisingly) thinks it is, and thus the study of consciousness is of less importance in the overall project of understanding the mind than we standardly think.

An approach is conscious-shy if it sidesteps or tries to finesse the study of consciousness for any of the above reasons. Conscious shyness has, as it were, four causes. A scientist of the mind can be conscious-shy because she is methodologically modest, practicing a sort of enlightened neopositivism. Alternatively, she might believe in the reality of consciousness but think that it names such a vague or, what is different, such a heterogeneous set of phenomena that it is not the sort of thing for which a general theory can be developed (at least not yet). Different still, she might be attracted to the brand of functionalism that thinks that the aim is to explain intelligent mentality in its most abstract form, that is, in terms of the metaphysically essential properties required for the explanation of intelligence in any possible world that contains it, not in terms of the contingent properties associated with just our kind of minds, which might, not surprisingly, be especially dear to our hearts. Finally, she might simply think that we should

postpone any general theorizing about the role and function of consciousness until the evidence gives further indication of the extent to which consciousness is involved in the causal fray, until, that is, there is evidence that it does play a significant role in the overall economy of mind. Even if consciousness plays a presidential role in the mind, there is still the question of whether it is presidential in the way Reagan was or the way FDR was.

Besides the causes cited so far for conscious shyness, there are still other causes for the even stronger reaction of eschewing altogether the scientific study of consciousness. Nonnaturalists have their own reasons for thinking that the problem of consciousness will not yield to science, but there are naturalists who think this too. Anticonstructive naturalism, or the new mysterianism, is the view that consciousness, despite being a natural phenomenon, will *never* be understood. Whether its causal role is significant or not, it will not be understood. The "old mysterians" were dualists who thought that consciousness cannot be understood scientifically because it operates according to nonnatural principles and possesses nonnatural properties. Consciousness might be understood in other ways, for example, by invoking supernatural phenomena (for some sophisticated contemporary defenses of supernaturalism, see Swinburne 1984 and Adams 1987). Unlike the old mysterianism or contemporary supernaturalism, the new mysterianism is a naturalistic position. Mind and consciousness exist, and they operate in accordance with natural principles and possess natural properties. But the new mysterianism is a postmodern position designed to drive a railroad spike through the heart of scientism, the view that science will eventually explain whatever is natural.

Colin McGinn is a new mysterian, a proponent of anticonstructive or noumenal naturalism. He thinks that naturalism must be true. There is no other credible way to think about the relation of consciousness and the brain than as a natural relation. Nonetheless, he thinks, we will never be able to set out a credible constructive theory of that relation.

McGinn writes, "We have been trying for a long time to solve the mind-body problem. It has stubbornly resisted our best efforts. The mystery persists. I think the time has come to admit candidly that we cannot resolve the mystery" (1989, 349). McGinn thinks that "we know that brains are the *de facto* causal basis of consciousness," but "we are cut off by our very cognitive constitution from achieving a conception of that natural property of the brain (or of consciousness) that accounts for the psychophysical link" (1989, 350). McGinn thinks that there is some natural property or set of properties *P* that accounts for consciousness. Consciousness comes with our kind of brain. It supervenes on neural activity in our kind of brain.

McGinn simply thinks that it is in principle impossible for us to comprehend *P*. (McGinn is somewhat unclear whether it is *P* that we cannot understand or *how* it is that *P* accounts for consciousness or makes the existence of consciousness “intelligible.”) What natural property *P* causes certain neural events to be consciously experienced by realizing these very conscious states is closed to us. If an omniscient God exists, then God knows all about *P* and how it gives rise to consciousness. But both *P* and the intelligible connection between *P* and consciousness are closed to us, the very systems in which the intelligible connections in fact obtain.

Although the doctrine is mischievous, coming from a naturalist, it is a coherent position. There are limitative results in physics and mathematics, for example, Heisenberg’s Uncertainty Principle and Gödel’s Incompleteness Theorem, that tell us of in-principle impossibilities faced by the physicist and mathematician. It is conceivable that just as we cannot know the position and momentum of an electron at one and the same time, or just as we can know that a certain sentence in arithmetic is true though it is in principle impossible for us to prove it within arithmetic, so we can know that consciousness is a natural phenomenon though it is in principle closed to us to know what sort of natural phenomenon it is.

It might be thought that there are limitative results within psychology itself that lend the position support. A child at each stage of cognitive development can understand certain questions or problems that are only solvable at the next stage, a stage she has not reached (Fodor 1991a). But the lesson of cognitive limits in mental development runs against the new mysterianism. For the child always eventually solves the problems she faces. None of the mysteries faced are terminal (Dennett 1991c).

It is important to see that the new mysterianism is different from principled agnosticism. The agnostic thinks that we do not understand what form a naturalistic solution to the consciousness-brain problem would take, so we ought not to confidently claim that naturalism is true. Thomas Nagel sometimes plays the agnostic. In his famous paper “What Is It Like to Be a Bat?” (1974), Nagel argues that there can be no remotely plausible naturalistic account of consciousness, that something essential will always be left out of even our very best theory. Nagel writes that “Consciousness is what makes the mind-body problem really intractable. . . . Without consciousness the mind-body problem would be much less interesting. With consciousness it seems hopeless” (1974, 165–166). But Nagel adds this important clarification: “It would be a mistake to conclude that physicalism is false. . . . It would be truer to say physicalism is a position we cannot understand because we do not at present have any conception of how it might be true” (1974, 176). In his book *The View from Nowhere*, Nagel puts it this way: “We have at present no conception of how a single event or thing could have both physical and phenomenological aspects, or how if it did they

might be related" (1986, 47). Because we do not understand what form a constructive naturalistic solution to the problem of consciousness would take, we cannot assign credibility to the claim that physicalism is true or to the claim that it is false. Intellectual honesty requires that we be agnostics.

Thanks to some recent work in neuropsychology, cognitive science, and philosophy, we have, I think, the beginnings of a credible theory of consciousness. We can therefore start to sketch the outlines of a constructive-naturalistic theory. Conscious shyness can be overcome, the new mysterianism can be shown to be at least premature, and perhaps some agnostics can be converted.

5 *Subjectivity, Objectivity, and the Natural Method*

The question is by what method consciousness is to be studied. I propose that we try the most natural strategy, what I call the *natural method*, to see if it can be made to work. Tactically, what I have in mind is this. Start by treating three different lines of analysis with equal respect. Give phenomenology its due. Listen carefully to what individuals have to say about how things seem. Also, let the psychologists and cognitive scientists have their say. Listen carefully to their descriptions about how mental life works and what jobs consciousness has, if any, in its overall economy. Finally, listen carefully to what the neuroscientists say about how conscious mental events of different sorts are realized, and examine the fit between their stories and the phenomenological and psychological stories.

The object of the natural method is to see whether and to what extent the three stories can be rendered coherent, meshed, and brought into reflective equilibrium. The only rule is to treat all three—the phenomenology, the psychology, and the neuroscience—with respect. Any a priori decision about which line of analysis “gets things right” or “has the last word” prejudices the question of whether different analyses might be legitimate for different explanatory purposes and thus compatible with each other, or at least capable of peaceful coexistence. As theory develops, analyses at each level are subject to refinement, revision, or rejection.

One might think that the natural method must either be the preferred method or at least have been tried (it is “natural,” after all) and have been shown to be deficient. This is a plausible but false thought. The natural method is certainly not the canonical method at the present time. Principled agnostics think that the concepts, categories, and methods of phenomenological, psychological, and neuroscientific analyses cannot be brought into harmonious relations. The new mysterianism or anticonstructive naturalism is the view that no method is promising. Any proposed methodology for cracking the mystery of consciousness will need to base its promises on the idle hope that the mystery of consciousness can be

made to yield. Some philosophers, Richard Rorty calls them "intuitive realists" (1982), favor phenomenology over all else. On the other side, verificationists, who are ruled by the law of silence on all matters not intersubjectively available in some strong sense, have terrible allergies to phenomenology. Meanwhile, many functionalists believe that neuroscience might yield an interesting picture about how mind is implemented, but they are skeptical that it can shed any light on how the mind works. Psychological explanation, according to standard functionalism, is autonomous. Eliminativist neuroscientists, on the other hand, think that the true and complete story of mind will emerge from the story of the brain and that those who start from higher-level descriptions are in constant danger of being led astray by the silly theories embedded in our ordinary ways of thinking about the mental. So, the natural method is decidedly not the currently preferred method.

The question remains, Was it tried and discarded because in fact it did not work? The answer is no. Phenomenology alone has been tried and tested. It does not work. But all we know is that taken *alone* it does not work. Phenomenology alone never reveals anything about how "seemings" are realized, nor can it reveal anything at all about the mental events and processes involved in conscious mental life, like acoustic or visual decoding, which don't *seem* any way at all. But it is incredible to think that we could do without phenomenology altogether.

Cognitive psychology alone also does not work. Explanations at the psychological level can provide illuminating models of mental activity. But psychological explanations need to be constrained by knowledge about the brain. There are always more functional hypotheses compatible with the facts than can be true. Functional psychological analyses involve posits of all manner of processes, processors, and processing. Neuroscience constrains such positing, but the study of the brain alone will yield absolutely no knowledge about the mind unless certain phenomena described at the psychological or phenomenological level are on the table to be explained.

Imagine an utterly complete explanation of the brain framed in the languages of physics, chemistry, and biochemistry. What has one learned about mental function? The answer is nothing, unless one's neuroscientific inquiry was guided in the first place by the attempt to map certain mental functions onto the brain. This is easy to see. There is much exciting work now underway linking robust phenomenological states with distinctive types of neural activity. For example, PET scans (positron emission tomography) reveal that normal anxiety, anticipatory fear, and panic attacks involve different kinds of brain activity (M. Raichle, personal communication). Such research not only draws us closer to understanding the deep structure of these different kinds of states, it also has important implications for psychopharmacology. It is well known that tranquilizers effective in

treating severe anxiety are not as effective in treating panic disorders. Sometimes antidepressants are more effective at preventing and moderating panic attacks than tranquilizers. It is natural to think that panic is simply extreme anxiety and thus simply more of whatever neural activity subserves anxiety. Despite certain strong phenomenological similarities, however, it appears that anxiety and panic involve profoundly different kinds of neural activity. The neurological difference is not one of degree but of kind.

6 *The Natural Method: Three Examples*

Let me give three very different examples how the natural method of seeking reflective equilibrium works.

Splitting auditory attention

There are many well-known studies of the splitting of auditory attention. In one well-known paradigm (Lackner and Garrett 1973), subjects are asked to pay attention only to the left channel in a set of earphones. In this channel they hear an ambiguous target sentence such as 'The lieutenant put out the lantern to signal the attack.' In the unattended right channel lots of irrelevant noise is produced plus the sentence 'He extinguished the flame.' Afterward the subjects are interviewed. This gives us a phenomenology. They tell us what they heard in the attended channel, and they insist that they heard nothing in the unattended channel. They were told, after all, not to listen. Subjects are good at keeping meaningful sounds received at the unattended channel from becoming conscious, or so it seems. This well-known phenomenon is known as the "Broadbent filtering effect."

It turns out, however, that when subjects are asked to choose between two interpretations of the target sentence 'The lieutenant put out the lantern to signal the attack'—(1) 'He extinguished the lantern to signal the attack' and (2) 'He put the lantern outdoors to signal the attack'—they display a decided preference for the interpretation consonant with the semantically related sentence 'He extinguished the flame', which they claim not to have heard in the unattended channel. It is known from controls that there is no preference for one meaning over the other without cuing. The best inference, therefore, is that the subjects did in fact "hear" the sentence that they are phenomenological blanks about.

There are several explanations at the psychological level compatible with these data. The favored explanation is that acoustical processing occurs in both the attended and unattended channels. The noise acoustically parsed in the unattended channel is sent on for semantic processing, and meaning is attached to the noise (how else can we explain the semantic effects of the "noise" in the unattended channel?). Meanwhile, the sentence

on the attended side is an object of explicit awareness, it is remembered, and it is available for explicit report afterward. The semantically decoded sentence on the unattended side is also remembered. But because the sentence was never consciously processed, it is not consciously retrievable from memory in the form in which it was received and processed. Nonetheless, because its semantic content was fixed and received by memory, this content affects the disambiguation task in a reliable way. Subjects are unaware of having heard the sentence 'He extinguished the flame', but the content of the sentence flows smoothly through the cognitive system and causes the subject to interpret 'put out' as 'extinguished'.

This explanation *could* be true. But how can we be assured that this psychological analysis is true? There are alternative analyses. Here is one alternative. What really happens is that the noise in the unattended channel is conscious for only an instant. The brevity of the conscious episode explains why it can't be remembered as it was in fact experienced.

The first analysis takes the phenomenology as authoritative about how things were in the experimental situation. The alternative analysis disregards this authority, at least partially. It seems to the subject as if he were never consciously aware of the noise in the unattended channel. But he is wrong. His memory fails him. At the instant the sentence "He extinguished the flame" came through the right channel, he heard it. There was a seeming connected to that noise. He just didn't remember it later.

How could one make a motivated choice between these two analyses of dichotic listening? Or, to consider a more radical alternative, how could one decide that both were worthless?

Here brain science might prove useful. Neuroscience will need to tell us whether, and if so, how, the functional distinctions made by both hypotheses among acoustic, semantic, conscious, and memory processing are preserved in the brain. The brain scientist cannot begin the study of linguistic processing in general and semantic disambiguation in particular, nor can she undertake to examine the legitimacy of the distinction between acoustic processing and semantic processing without taking both the phenomenological and functional accounts seriously. But whether the distinction between acoustic and semantic processing can survive careful study of the process of linguistic comprehension by brain scientists remains to be seen.

Suppose that the psychologists' functional distinctions among acoustical and semantic processors, memory, and the like remain intact and that we are left with the choice between the original hypothesis involving no consciousness of the contents in the unattended channel and the alternative hypothesis of a lost flash of consciousness. There are philosophers who think that brain science will never be able to help us choose between hypotheses at this level of grain (Dennett 1991a). But I say, "Never say

never." How conscious mental states are realized is something that neuroscience may someday yield fine-grained information about. It has recently been suggested that subjective awareness is linked to oscillation patterns in the 40 hertz range in the relevant groups of neurons; that is, neurons involved in a certain decoding task "synchronize their spikes in 40 Hz oscillations" (Crick and Koch 1990, 272). The 40 hertz patterns can be sustained for very short periods of time, in which case there is rapid memory decay, or they can resonate for several seconds, in which case they become part of working memory, give rise to more vivid phenomenology, and are more memorable. Suppose that this hypothesis turns out to be corroborated across sensory modalities and that short-term 40-hertz oscillations are observed to occur when the sentence in the unattended channel is presented. Combining present theories of short-term and working memory with such a finding would lend support to the second, initially weird, idea that the sentence in the unattended channel makes a conscious appearance (it's a 40-hertz oscillation, after all) but it is not remembered.

Neural correlates of subjective visual attention

One way to study conscious experience is to correlate particular qualitative types of experience with particular kinds of activity. One fascinating set of studies linking particular types of awareness with particular types of neural activity has recently been done on rhesus macaques and exploits the well-known phenomenon of rivalry. Gestalt images, like the necker cube or the duck-rabbit illusion, are rivalrous. The stimulus pattern stays the same, but perception flip-flops. Binocular rivalry is a particular type of rivalry that exploits the fact that the visual system tries to come up with a single percept even though the eyes often receive different visual information. In experiments involving binocular rivalry the visual input presented to the two eyes is incompatible. For example, a line moving upward is presented on the left, and a line moving downward is presented on the right. "Because such stimuli cannot be fused by the cyclopean visual system, the perception alternates between the right eye alone or the left eye alone" (Logothetis and Schall 1989, 761). Humans report such alternations in perception. But how, one might ask, does one find out a monkey's phenomenology? The answer is that one trains the monkey prior to the experiment to be a reliable reporter of whether it perceives a line moving up or a line moving down. This can be done by training monkeys to give bar-press reports or, more surprisingly, by training the monkeys to execute a saccade (a quick eye movement) to a spot on the left if a downward movement is perceived and to a spot on the right if an upward movement is perceived. A monkey's report of how things appear at any moment (its phenomenology) provides psychophysical data about the rate of perceptual shifting and raises interesting questions about why there is shifting

perception as opposed to a winner-take-all lock on one of the rival perceptual interpretations. But the phenomenological data ("The line is now moving upward") and the psychophysical data about the time between perceptual switches yield no information about what is going on in the brain, and in particular they yield no information about what neuronal events are involved in the shifting perceptions. This is where looking at the brain helps.

It is well known that monkeys have as many as 20 to 25 visual areas in the neocortex. The bulk of retinal output projects through the lateral geniculate nucleus of the thalamus to the primary visual cortex at the back of the brain that computes edges. Other areas are interested in color and shape, position relative to the organism itself, facial features, and motion and depth. Motion detection occurs primarily in the middle temporal and medial superior temporal areas in the superior temporal sulcus (STS). So the activity of neurons in the STS were monitored as the monkeys reported upward or downward motion. The study of the effect of the rival stimuli on 66 single neurons indicated that the "activity of many neurons was dictated by the retinal stimulation. Other neurons, however, reflected the monkeys' reported perception of motion direction" (Logothetis and Schall 1989, 761). The principle of supervenience says that every difference at the level of mentality must be subserved by a difference at some lower level (but perhaps not conversely). This experiment indicates how the robust phenomenological difference between an upward or downward moving image might be subserved (although undoubtedly not exclusively) by small but detectable changes in activity in the cortical areas subserving motion detection.

The experiment is an excellent example of how subjective awareness can be studied by drawing together information gathered at different levels of analysis and by distinctive techniques. First, there is the assumption that there is something it is like for a monkey to have visual experience. Second, good old-fashioned psychological techniques of operant conditioning are used to train the monkeys to provide reports about what they see. Finally, these reports are linked with detailed observations of the activity of 66 distinct neurons to yield information about the distinct brain processes subserving perceptual experiences of upward and downward moving lines. This is knowledge by triangulation. The natural method works.

Conscious event memory

Many years ago a famous neurological patient H.M. had medial-temporal-lobe surgery to cure his epilepsy. H.M.'s ability to remember new events and facts died with the excision. Still, H.M. remembered how to perform standard motor tasks and how to use language. Indeed, to this day H.M. is something of a crossword puzzle aficionado. His intact semantic memory is

tapped by questions directly before his eyes, and his place in the puzzle is visually available to him. When H.M. is away from a puzzle he is working on, he cannot remember how far along on it he is or even that he is working on a puzzle. There are other sorts of games or puzzles that H.M. was not familiar with before surgery but that he has become good at. The Tower of Hanoi puzzle requires that one move a set of rings down a three-post line-up (one ring and one post at a time) so that in the end they are in order of size on the last post. When the game was first taught to H.M., he played as poorly as any first-timer. But gradually he caught on. Each time the puzzle is presented to him, he claims never to have seen it before. But even after long periods of not working the puzzle, H.M. shows the clear effects of prior experience. He is better than any novice and invariably starts at a level of proficiency close to where he left off. H.M. has no conscious declarative memories about the Tower of Hanoi puzzle, but quite clearly information has sunk in about how to do it.

H.M.'s good general intelligence and his semantic memory can sustain him during a short period of newspaper reading. But minutes after he has read the newspaper, he is blank about what he read. It is tempting to say that H.M. has an intact short-term memory but no long-term memory. But this is too simplistic for a variety of reasons. First, he retains information over the long term about how to do the Tower of Hanoi puzzle. The information is just not retained consciously. Second, when he does a crossword puzzle or reads a newspaper, he depends on something longer than short-term memory, which on standard views is very short. As long as H.M. is paying attention to a task, like a crossword puzzle or a newspaper article, he can retain information about it. When his attention is removed from the overall task his memory goes blank.

H.M. can give us a phenomenology, a set of reports about what he remembers and what he doesn't. He also reveals in his behavior that he remembers certain things that he can't consciously access. Other human amnesiacs provide further phenomenological and behavioral data. Putting together the phenomenological and behavioral data with the knowledge that there is damage or excision to the medial temporal lobe leads to an initial hypothesis that this area plays some important role in fixation and retrieval of memories of conscious events.

The medial temporal lobe is a large region that includes the hippocampus and associated areas, as well as the amygdala and related areas. Magnetic resonance imaging (MRI) has allowed for very precise specification of the damaged areas in living human amnesiacs and in monkeys. This research reveals that the hippocampus is the crucial component. When there is serious damage or removal of the hippocampal formation, the entorhinal cortex, and the adjacent anatomically related structures, the perirhinal and parahippocampal cortices, the ability to consciously remem-

ber novel facts or events is lost. Removal or serious lesions of the amygdala profoundly affect emotions, but not memory. It is not as if memories were created and set down in the hippocampal formation. The hippocampal formation is necessary to lay down memories, but it is not remotely sufficient for the conscious memory of facts and events. For habit and skill learning, it is not even necessary.

The complex network involved in fixing such memories has now been mapped in some detail by using monkey models and by drawing inferences about human memory function based on comparative deficit and normal data. Larry Squire and Stuart Zola-Morgan (1991) report an elegant experiment in which monkeys were given a simple recall task. Each monkey was presented with a single object. Then after a delay (15 seconds, 60 seconds, 10 minutes), it was presented with a pair consisting of the original object and a new one. The monkeys were trained to pick the novel object to get a reward. The task is trivial for monkeys with intact hippocampal formations, but there is severe memory impairment (increasing with latency) in monkeys with destruction to the hippocampal formation. The procedure is a reliable test of perceptual memory.

A skeptic could admit this but deny that there is any evidence that monkeys *consciously* remember what they saw. To be sure, the information about the original stimulus is processed and stored. The monkeys' behavior shows that these memories are laid down. But for all we know, these memories might be completely unconscious, like H.M.'s memories for the Tower of Hanoi puzzle. They might be. However, on the basis of anatomical similarities between monkey and human brains, the similarity of memory function, and evolutionary considerations, it is credible that the monkeys' selection of the appropriate stimulus indicates what they consciously remember, and thus that it can be read as providing us with a phenomenology, with a set of reports of how things seem to them.

The next step is to put the phenomenological and behavioral data from humans and monkeys together with the data about specific types of lesions suffered and join both to our best overall theories of the neural functions subserving memory and perception. Coordination of these theories and data yield the following general hypothesis. When a stimulus is observed, the neocortical area known to be sensitive to different aspects of a visual stimulus are active. For example, in the simple matching task, the areas responsible for shape and color detection are active. In a task where the object is moving, the areas involved in motion detection become involved. Activity of the relevant cortical areas is sufficient for perception and immediate memory. "Coordinated and distributed activity in neocortex is thought to underlie perception and immediate (short-term) memory. These capacities are unaffected by medial temporal lobe damage.... As long as a percept is in view or in mind, its representation remains coherent in short-

term memory by virtue of mechanisms intrinsic to neocortex" (Squire and Zola-Morgan 1991, 1384). A memory is set down only if the information at the distributed cortical sites is passed to three different areas close to the hippocampus and then into the hippocampus itself. The hippocampus then passes the message back through the medial temporal lobe out to the various originating sites in the neocortex that processed the relevant aspects of the stimulus in the first place. The memory is laid down in a distributed fashion. It is activated when the connections between the hippocampus and the various areas it projects to are strengthened by, for example, the request for recall. Once the memory is laid down and especially after it is strengthened, "Proust's principle" comes into play. The memory can be reactivated by the activation of any important node, e.g., the one subserving object shape or color, even smell or sound, without having to be turned on by the hippocampus directly. In this way the hippocampal formation passes the "burden of long-term (permanent) memory storage" to the neocortex and is freed to make new memories (Squire and Zola-Morgan 1991, 1385). This theory is credible and powerful. It explains how damage to the hippocampal formation can destroy the capacity to form new memories, and it explains why old memories are retained despite hippocampal destruction—their storage has been assumed by the neocortex.

Here again is an elegant example of the natural method. There are the phenomenological and behavioral data coming from H.M. and other human patients. There are the behavioral data coming from the monkeys, which I claim ought to be read as informing us about how things seem to the monkeys. There is the prior work linking visual processing with certain brain processes. And there is the eventual theory that explains how conscious memories are fixed, how they are recalled, what they are doing when they are not active (they are dispositions laid down in neural nets), and so on.

The three examples taken together show the power of the natural method. They also show that there is nothing about the phenomenological concept of consciousness that renders it logically incompatible with the terms and concepts of a science of the mind. The best strategy is one that promotes the search for relations among the phenomenological, information-processing, and neural levels (Wimsatt 1976; P. S. Churchland 1986; Shallice 1988; McCauley, in press). In practice, the search for relations will require give-and-take in various places, especially in our commonsense characterizations of mental events. There are a host of possible relations that might be discovered to obtain among the different levels. Presumably, each and every mental state is a neural state; that is, presumably, token physicalism is true. But there are multifarious relations of dependency,

covariation, and supervenience that might obtain between mental events and neural events (see Kim 1990 for an analysis of many of these relations). The evidence suggests that there are some domains, for example, color vision, in which relatively neat mappings exist between mental-state types and neural-state types and thus for which some sort of identity theory may be true. There are other domains, for example, semantic memory, where strict type identity between the mental state that subserves my understanding of the term 'chair' and the mental state that subserves yours is less credible (of course, this may well be due to the fact that despite *prima facie* semantic sameness, 'chair' for me and 'chair' for you do not belong to the exact same mental kind, since they mean somewhat different things). There may well be concepts and theories designed for understanding phenomena at the macro level, concepts such as that of the self and theories of self-consciousness and narrative self-representation, for example, that are autonomous in the sense that they do not depend on showing how self-representation is achieved at the micro level in the brain (Fodor 1981; McCauley 1986, in press). The science of the mind requires an account of self-consciousness and self-representation, since these are genuine phenomena. The account must be compatible with what is deemed possible and credible by neuroscience, even if our theories of the self are not perspicuously expressible in neuroscientific terms. In the end, the best bet is that all the various relations mapped out as possibilities in the debates among identity theorists, functionalists, and eliminativists—relations of type identity, token identity, reduction, autonomy, and outright elimination (perhaps there is no id)—obtain in some domains.

The method of seeking reflective equilibrium among the phenomenological, psychological, and neuroscientific levels is deployed throughout this essay. Its merits are for others to judge. But if I am right that there are insights about the conscious mind to be gained at the level of first-person and third-person phenomenology, at the level of functional psychology, and at the level of brain science, it would be foolish to proceed in any other way.

References

-
- Adams, Robert M. 1987. "Flavors, Colors, and God." In *The Virtue of Faith and Other Essays in Philosophical Theology*. New York: Oxford University Press.
- Baars, B. J. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Bandura, A. 1986. *The Social Foundation of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, N.J.: Prentice-Hall.
- Block, N., ed. 1980a. *Readings in Philosophy of Psychology*. 2 vols. Cambridge: Harvard University Press.
- Block, N. 1980b. "Troubles with Functionalism." In Block 1980a.
- Block, N. 1991. "What Does Neuropsychology Tell Us about a Function of Consciousness?" Unpub. ms.
- Block, N. 1992. "Begging the Question against Phenomenal Consciousness." *Behavioral and Brain Sciences*. Comment on Dennett and Kinsbourne 1992.
- Bradford, H. F. 1987. "Neurotransmitters and Neuromodulators." In Gregory 1987.
- Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press.
- Bruner, J. S. 1983. *In Search of Mind: Essays in Autobiography*. New York: Harper and Row.
- Bruner, J. S. 1986. *Actual Minds, Possible Worlds*. Cambridge: Harvard University Press.
- Calvin, W. H. 1990. *The Cerebral Symphony: Seashore Reflections on the Structure of Consciousness*. New York: Bantam.
- Calvin, W. H. 1991. *The Ascent of Mind: Ice Age Climate and the Evolution of Intelligence*. New York: Bantam.
- Changeux, J. P. 1985. *Neuronal Man: The Biology of Mind*. Translated by L. Garey. Oxford: Oxford University Press.
- Churchland, Patricia S. 1983. "Consciousness: The Transmutation of a Concept." *Pacific Philosophical Quarterly* 64:80–93.
- Churchland, Patricia S. 1986. *Neurophilosophy*. Cambridge: MIT Press.
- Churchland, Patricia S. 1988. "Reduction and the Neurobiological Basis of Consciousness." In Marcel and Bisiach 1988.
- Churchland, Paul M. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78:67–90.
- Churchland, Paul M. 1988. *Matter and Consciousness*. Rev. ed. Cambridge: MIT Press.
- Churchland, Paul M. 1989. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge: MIT Press.
- Churchland, Paul M., and Patricia S. Churchland. 1990. "Intertheoretic Reduction: A Neuroscientist's Field Guide." *Seminars in the Neurosciences* 2:249–256.
- Craik, K. 1943. *The Nature of Explanation*. Cambridge: Cambridge University Press.
- Crick, Francis, and Christof Koch. 1990. "Towards a Neurobiological Theory of Consciousness." *Seminars in the Neurosciences* 2:263–275.

- Dawkins, R. 1976. *The Selfish Gene*. Oxford: Oxford University Press.
- Dennett, D. 1969. *Content and Consciousness*. London: Routledge and Kegan Paul.
- Dennett, D. 1971. "Intentional Systems." In Dennett 1978.
- Dennett, D. 1978. *Brainstorms*. Cambridge: MIT Press.
- Dennett, D. 1982. "How to Study Consciousness Empirically; or, Nothing Comes to Mind." *Synthese* 59:159–180.
- Dennett, D. 1986. "Julian Jaynes's Software Archeology." *Canadian Psychology* 27, no. 2:149–154.
- Dennett, D. 1988a. "Why Everyone Is a Novelist." *Times Literary Supplement* 4, no. 459 (September): 1016–1022.
- Dennett, D. 1988b. "The Evolution of Consciousness." CCM-88-1. Center for Cognitive Studies, Tufts University.
- Dennett, D. 1988c. "Quining Qualia." In Marcel and Bisiach 1988.
- Dennett, D. 1989. "The Origins of Selves." *Cogito* 2:163–173.
- Dennett, D. 1991a. *Consciousness Explained*. New York: Little Brown.
- Dennett, D. 1991b. Review of Colin McGinn, *The Problem of Consciousness*. *Times Literary Supplement*, May 10.
- Dennett, D. 1991c. "The Problem of Consciousness." *Times Literary Supplement*, Letter. June 14.
- Dennett, D. 1991d. "Real Patterns." *Journal of Philosophy* 88:27–51.
- Dennett, D., and M. Kinsbourne. 1992. "Time and the Observer: The Where and When of Consciousness in the Brain." *Behavioral and Brain Sciences* 15:183–247.
- Dewey, John. 1922. *Human Nature and Conduct*. New York: Henry Holt, 1957.
- Dixon, N. F. 1987. "Subliminal Perception." In Gregory 1987.
- Edelman, G. M. 1987. *Neural Darwinism*. New York: Basic Books.
- Edelman, G. M. 1989. *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.
- Ekman, P., R. W. Levinson, W. V. Freisen. 1985. "Autonomic Nervous-System Activity Distinguishes among Emotions." *Science* 221:1208–1210.
- Erikson, Erik H. 1968. *Identity: Youth and Crisis*. New York: W. W. Norton.
- Eysenck, H. J. 1967. *The Biological Basis of Personality*. Springfield, Ill.: Charles C. Thomas.
- Fehrer, E., and D. Raab. 1962. "Reaction Time to Stimuli Masked by Metacontrast." *Journal of Experimental Psychology* 63, no. 2:143–147.
- Flanagan, O. 1991a. *Varieties of Moral Personality: Ethics and Psychological Realism*. Cambridge: Harvard University Press.
- Flanagan, O. 1991b. *The Science of the Mind*. 2nd ed. MIT Press.
- Fodor, J. 1975. *The Language of Thought*. New York: Crowell.
- Fodor, J. 1981. "Special Sciences." In *Representations*. Cambridge: MIT Press.
- Fodor, J. 1991a. "Too Hard for Our Kind of Mind?" Review of Colin McGinn, *The Problem of Consciousness*. *London Review of Books*, p. 12.
- Fodor, J. 1991b. "The Problem of Consciousness." Letter. *Times Literary Supplement*, June 7.
- Gardner, Howard. 1983. *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Gazzaniga, M. S. 1988. *Mind Matters: How Mind and Brain Interact to Create Our Conscious Lives*. Boston: Houghton Mifflin.
- Goldman, A. I. 1986. *Epistemology and Cognition*. Cambridge: Harvard University Press.
- Goldman, A. I. 1991. "The Psychology of Folk Psychology." Paper given at the annual meetings of the Society for Philosophy and Psychology, San Francisco.
- Gould, S. J., and R. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London*, series B, 205.

- Gregory, R., ed. 1987. *The Oxford Companion to Mind*. Oxford: Oxford University Press.
- Griffin, Donald R. 1981. *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience*. Los Altos, Calif.: William Kaufmann.
- Hacking, Ian. 1991. "Two Souls in One Body." *Critical Inquiry* 17, no. 4:838–867.
- Hardin, C. L. 1988. *Color for Philosophers: Unweaving the Rainbow*. Indianapolis: Hackett.
- Henriques, J. B., and R. J. Davidson. 1990. "Regional Brain Electrical Asymmetries Discriminate between Previously Depressed and Healthy Control Subjects." *Journal of Abnormal Psychology* 99, no.1: 22–31.
- Hobson, J. Allan. 1988. *The Dreaming Brain*. New York: Basic Books.
- Hume, David. 1777. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Ed. L. A. Selby-Bigge. Oxford: Oxford University Press, 1975.
- Humphrey, N., and D. C. Dennett. 1989. "Speaking for Ourselves." *Raritan: A Quarterly Review* 9:69–98.
- Hundert, E. 1989. *Philosophy, Psychiatry, and Neuroscience: Three Approaches to Mind*. New York: Oxford University Press.
- Jackendoff, R. 1987. *Consciousness and the Computational Mind*. Cambridge: MIT Press.
- Jackson, F. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32:127–136.
- Jackson, F. 1986. "What Mary Didn't Know." *Journal of Philosophy* 83, no. 5:291–295. Reprinted in D. Rosenthal, ed., *The Nature of Mind*. New York: Oxford University Press, 1991.
- James, William. 1890. *The Principles of Psychology*. 2 vols. New York: Dover, 1950.
- James, William. 1892. *Psychology: The Briefer Course*. Ed. G. Allport. New York: Harper and Row, 1961.
- Jaynes, J. 1976. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Boston: Houghton Mifflin.
- Johnson, Mark. 1987. *The Body in the Mind*. Chicago: University of Chicago Press.
- Johnson-Laird, P. N. 1983. *Mental Models*. Cambridge: Harvard University Press.
- Johnson-Laird, P. N. 1988. *The Computer and the Mind: An Invitation to Cognitive Science*. Cambridge: Harvard University Press.
- Kagan, J. 1984. *The Nature of the Child*. New York: Basic Books.
- Kagan, J. 1989. *Unstable Ideas: Temperament, Cognition, and Self*. Cambridge: Harvard University Press.
- Kermode, Frank. 1967. *The Sense of an Ending: Studies in the Theory of Fiction*. New York: Oxford University Press.
- Kim, Jaegwon. 1990. "Supervenience as a Philosophical Concept," C-3, no. 23, Research Group on Mind and Brain, University of Bielefeld.
- Lackner, J., and M. Garrett. 1973. "Resolving Ambiguity: Effects of Biasing Context in the Unattended Ear." *Cognition* 1:359–372.
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Levine, J. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64:354–361.
- Lewis, David. 1980. "Mad Pain and Martian Pain." In Block 1980a.
- Libet, B. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8:529–566.
- Lloyd, Dan. 1989. *Simple Minds*. Cambridge: MIT Press.
- Loar, Brian. 1990. "Phenomenal Properties." In *Philosophical Perspectives: Action Theory and Philosophy of Mind*. Atascadero, Calif.: Ridgeview.
- Locke, John. 1690. *An Essay Concerning Human Understanding*. Ed. P. H. Nidditch. 2nd ed., 1694. 5th ed., 1706. Oxford: Oxford University Press, 1975.
- Logothetis, N., and J. D. Schall. 1989. "Neuronal Correlates of Subjective Visual Perception." *Science* 245:761–763.

- Luria, A. R. 1972. *The Man with the Shattered World*. Cambridge: Harvard University Press, 1987.
- Lycan, W. 1981. "Form, Function, and Feel." *Journal of Philosophy* 78:23–49.
- Lycan, W. 1987. *Consciousness*. Cambridge: MIT Press.
- McCauley, R. N. 1986. "Intertheoretic Reduction and the Future of Psychology." *Philosophy of Science* 53:179–199.
- McCauley, R. N. In press. "Cross-Scientific Study and the Complexity of Psychology." *Annals of Theoretical Psychology*, vol. 8.
- McGinn, C. 1991. *The Problem of Consciousness*. Oxford: Blackwell.
- MacIntyre, A. 1981. *After Virtue*. London: Duckworth.
- Mandler, George. 1985. *Cognitive Psychology: An Essay in Cognitive Science*. New York: L. Erlbaum.
- Marcel, A. J. 1988. "Phenomenal Experience and Functionalism." In Marcel and Bisiach 1988.
- Marcel, A. J., and E. Bisiach, eds. 1988. *Consciousness in Contemporary Science*. Oxford: Oxford University Press.
- Marr, David. 1982. *Vision*. San Francisco: Freeman.
- Minsky, M. 1985. *The Society of Mind*. New York: Simon and Schuster.
- Myers, Gerald E. 1986. *William James: His Life and Thought*. New Haven: Yale University Press.
- Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" In *Mortal Questions*. Cambridge: Cambridge University Press, 1979.
- Nagel, Thomas. 1986. *The View from Nowhere*. New York: Oxford University Press.
- Peirce, C. S. 1898. *Collected Papers*. Vol. 6, *Scientific Metaphysics*. Ed. C. Hartshorne and P. Weiss. Cambridge: Harvard University Press, 1935.
- Phelps, M. E., and J. C. Mazziotta. 1985. Positron Emission Tomography: Human Brain Function and Biochemistry. *Science* 228:799–809.
- Popper, K., and J. Eccles. 1977. *The Self and Its Brain*. New York: Springer-Verlag.
- Quine, W. V. O. 1952. "On Mental Entities." In *The Ways of Paradox and Other Essays*. New York: Random House, 1966.
- Quine, W. V. O. 1985. "States of Mind." *Journal of Philosophy* 82, no. 1:5–8.
- Raichle, Marcus E. 1987. "Images of the Brain in Action." In Gregory 1987.
- Rey, Georges. 1983. "A Reason for Doubting the Existence of Consciousness." In *Consciousness and Self-Regulation*, vol. 3, ed. R. J. Davidson, G. E. Schwartz, and D. Shapiro. New York: Plenum.
- Rey, Georges. 1988. "A Question about Consciousness." In *Perspectives on Mind*, ed. H. R. Otto and J. A. Tuedio. Dordrecht: Reidel.
- Rorty, R. 1982. *Consequences of Pragmatism*. Minneapolis: University of Minnesota Press.
- Rosenthal, David. 1986. "Two Concepts of Consciousness." *Philosophical Studies* 49:329–359.
- Rosenthal, David. 1989. "Thinking That One Thinks." ZIF report no. 11, Research Group on Mind and Brain, University of Bielefeld, Germany.
- Rosenthal, David. 1990. "Why Are Verbally Expressed Thoughts Conscious?" ZIF report no. 32, Research Group on Mind and Brain, University of Bielefeld, Germany.
- Rumelhart, D., J. McClelland, and the PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. 2 vols. Cambridge: MIT Press. Order of authors is McClelland and Rumelhart for vol. 2.
- Ryle, Gilbert. 1949. *The Concept of Mind*. London: Hutchinson.
- Sacks, Oliver. 1983. *Awakenings*. New York: Dutton.

- Sacks, Oliver. 1985. *The Man Who Mistook His Wife for a Hat and Other Clinical Tales*. New York: Summit.
- Sandel, M. 1982. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Schacter, D. 1989. "On the Relation between Memory and Consciousness: Dissociable Interactions and Conscious Experience." In *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, ed. Henry L. Roediger III and F. I. M. Craik. Hillsdale, N.J.: L. Erlbaum.
- Searle, J. 1984. *Minds, Brains, and Science*. Cambridge: Harvard University Press.
- Shallice, T. 1988. *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.
- Shweder, Richard. 1991. *Thinking through Cultures*. Cambridge: Harvard University Press.
- Smolensky, Paul. 1988. "On the Proper Treatment of Connectionism." *Behavioral and Brain Sciences* 11:1–23.
- Spence, D. 1982. *Narrative Truth and Historical Truth: Meaning and Interpretation in Psychoanalysis*. New York: W. W. Norton.
- Squire, L., and S. Zola-Morgan. 1991. "The Medial Temporal Lobe Memory System." *Science* 253:1380–1386.
- Stern, D. N. 1985. *The Interpersonal World of the Infant*. New York: Basic Books.
- Stich, S. 1978. "Autonomous Psychology and the Belief-Desire Thesis." *The Monist*, vol. 61.
- Swinburne, R. 1984. "Personal Identity: The Dualist Theory." In *Personal Identity*, ed. S. Shoemaker and Richard Swinburne. Oxford: Blackwell.
- Taylor, Charles. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge: Harvard University Press.
- Treisman, A. M., and H. Schmidt. 1982. Illusory Conjunctions in the Perceptions of Objects. *Cognitive Psychology* 14:107–141.
- Van Gulick, R. 1988. "A Functionalist Plea for Self-Consciousness." *Philosophical Review* 97, no. 2: 149–181.
- Van Gulick, R. 1990. "What Difference Does Consciousness Make?" *Philosophical Topics* 17:211–230.
- Van Gulick, R. In press a. "Understanding the Phenomenal Mind: Are We All Just Armadillos?" In *Consciousness: A Mind and Language Reader*, ed. M. Davies and G. Humphrey. Oxford: Blackwell.
- Van Gulick, R. In press b. "Consciousness May Still Have a Processing Role to Play." *Brain and Behavioral Sciences*.
- Varela, F. J., E. Thompson, E. Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.
- Velmans, M. In press. "Is Human Information Processing Conscious?" *Brain and Behavioral Sciences*.
- Von der Heydt, R., E. Peterhans, and G. Baumgartner. 1984. "Illusory Contours and Cortical Neuron Response." *Science* 224:1260–1262.
- Weiskrantz, L. 1988. "Some Contributions of Neuropsychology of Vision and Memory to the Problem of Consciousness." In Marcel and Bisiach 1988.
- Weiss, Rick. 1990. "A Flight of Fancy Mathematics: Chaos Brings Harmony to a Birder's Puzzle." *Science News* 137:172.
- Wellman, Henry M. 1990. *The Child's Theory of Mind*. Cambridge: MIT Press.
- Wimsatt, W. 1976. "Reductionism, Levels of Organization, and the Mind-Body Problem." In *Consciousness and the Brain*, ed. G. Globus, G. Maxwell, and I. Savodnik. New York: Plenum.
- Wilkes, K. V. 1988a. *Real People: Personal Identity without Thought Experiments*. Oxford: Oxford University Press.
- Wilkes, K. V. 1988b. "_____, Yishi, Duh, Um, and Consciousness." In Marcel and Bisiach 1988.