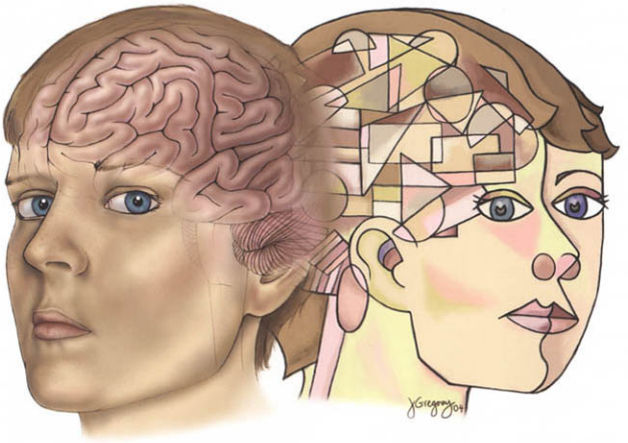
THE LOST SELF

PATHOLOGIES OF THE BRAIN AND IDENTITY



Edited by

TODD E. FEINBERG

JULIAN PAUL KEENAN

The Lost Self

This page intentionally left blank

The Lost Self

Pathologies of the Brain and Identity

Edited by

TODD E. FEINBERG
JULIAN PAUL KEENAN

OXFORD
UNIVERSITY PRESS
2005

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016
www.oup.com

Oxford is a registered trademark of Oxford University Press

Library of Congress Cataloging-in-Publication Data

The lost self : pathologies of the brain and identity / edited by
Todd E. Feinberg, Julian Paul Keenan.
p. ; cm.
Includes bibliographical references and index.
ISBN-13 978-0-19-517341-3
ISBN 0-19-517341-4
1. Depersonalization. 2. Self. 3. Identity (Psychology) 4. Cognitive
neuroscience. [DNLM: 1. Brain—physiopathology. 2. Ego.
3. Mental Disorders—physiopathology. WL 103.5 L881 2005]
I. Feinberg, Todd E. II. Keenan, Julian Paul.
RC553.D4L67 2005
616.8 — dc22 2004024803

2 4 6 8 9 7 5 3 1

Printed in the United States of America
on acid-free paper

Preface

I originally came to the study of the self as a clinician. In the course of my training and practice I had the opportunity to examine an array of interesting patients who had disordered selves as a result of brain pathology. For instance, I could not help but be struck by the oddness of a female patient of mine who suffered a stroke and then sang pop tunes to her paralyzed arm in the hope of bringing it back to life, or the woman who tried to throw her similarly motionless limb in the garbage in her hospital room. Some patients I witnessed screaming at their reflections in a mirror; others told tales of imaginary alter egos, or grappled with one of their arms as it tried to strangle them. Over the last 20 years the core of my work has involved the study of these cases and their disorders.

Three years ago I met Julian Keenan, co-editor of this volume, who is one of a growing breed of neuroscientists engaged in the experimental study of the self. Dr. Keenan has employed diverse imaging and nonimaging experimental methods in examining the brain correlates of the self in normal subjects. Together, in this volume we bring to the reader contributions from an eclectic group of original thinkers who explore the current state of our knowledge of the philosophical, neuropsychological, and neurobiological basis of the self and in particular how the self is transformed by brain pathology.

As I read these chapters, I was struck by how far the scientific study of the self has come. As is often—if not always—the case in science, it is also clear how far we have to go, but I hope the reader will agree that the journey has begun.

Many people in many roles have helped with this project. We especially thank Fiona Stevens, our editor at Oxford, for helping with the development of this project from its inception. Her expert guidance, advice, and support were invaluable. Thanks are due as well to Edith Barry. We also thank Jill Gregory for her beautiful cover art and her help with the figures in several chapters.

This page intentionally left blank

Contents

Contributors

SIMON BARON-COHEN, PH.D.
Autism Research Centre
Cambridge University
Department of Psychology and Psychiatry
Cambridge, United Kingdom

SARAH-JAYNE BLAKEMORE, PH.D.
Royal Society Dorothy Hodgkin Research Fellow
Institute of Cognitive Neuroscience
University College London
London, United Kingdom

WILLIAM CHRISTIANA, B.A.
Cognitive Neuroimaging Laboratory
Department of Psychology
Montclair State University
Upper Montclair, NJ

JOHN DELUCA, PH.D., ABPP
Director of Neuroscience Research
Kessler Medical Rehabilitation Research and Education Corporation
West Orange, NJ and
Professor of Physical Medicine and Rehabilitation, and of Neurosciences
UMDNJ-New Jersey Medical School
Newark, NJ

MARTHA J. FARAH, PH.D.
Bob and Arlene Kogod Term Professor of Psychology
Department of Psychology
University of Pennsylvania
Director Center for Cognitive Neuroscience
Philadelphia, PA

TODD E. FEINBERG, M.D.
Professor of Clinical Psychiatry and Behavioral Sciences and Clinical Neurology
Albert Einstein College of Medicine
Chief, Yarmon Neurobehavior and Alzheimer's Disease Center
Beth Israel Medical Center
New York, NY

ESTHER FUJIWARA, PH.D.
The Rotman Research Institute
Baycrest Centre for Geriatric Care
Toronto, Ontario, Canada

JOSEPH T. GIACINO, PH.D
Associate Director of Psychology
JFK Medical Center
Center for Head Injuries
Edison, NJ and
Department of Neuroscience
Seton Hall University
South Orange, NJ

SETH GILLIHAN, M.A.
Department of Psychology
University of Pennsylvania
Philadelphia, PA

GEORG GOLDENBERG, M.D.
Neuropsychologische Abteilung
Krankenhaus München Bogenhause
München, Germany

KEVIN GUISE
Cognitive Neuroimaging Laboratory
Department of Psychology
Montclair State University
Upper Montclair, NJ

J. ALLAN HOBSON, M.D.
Professor of Psychiatry
Harvard Medical School
Laboratory of Neurophysiology
Massachusetts Mental Health Center
Brookline, MA

JULIAN PAUL KEENAN, PH.D.
Director, Cognitive Neuroimaging Laboratory
Assistant Professor
Department of Psychology
Montclair State University
Upper Montclair, NJ

TROELS W. KJAER, M.D.
Clinical Neurophysiology Clinic
Copenhagen University Hospital
Copenhagen, Denmark

HEDY KOBER, B.A.
Department of Psychology
Columbia University
New York, NY

HANS C. LOU, M.D.
Department of Functionally Integrative Neuroscience
Aarhus University Hospital
Aarhus, Denmark

SARAH MALCOLM, B.A.
Cognitive Neuroimaging Laboratory
Department of Psychology
Montclair State University
Upper Montclair, NJ

HANS J. MARKOWITSCH, PH.D.
University of Bielefeld
Department of Physiological Psychology
Bielefeld, Germany

ROY J. MATHEW, M.D.
Professor of Medicine-Psychiatry
Department of Internal Medicine
Texas Tech Health Sciences Center
Odessa, TX

BRUCE L. MILLER, M.D.
A. W. & Mary Margaret Clausen Distinguished Professor of Neurology
Director, UCSF Memory & Aging Center and Alzheimer Disease
Research Center
San Francisco, CA

SUKHVINDER OBHI, PH.D.
Department of Psychology
University of Western Ontario
London, Ontario
Canada

KAREN SPANGENBERG POSTAL, PH.D., ABPP-CN
Neuropsychology Consultants
Andover, MA

ALYSA RAY, B.A.
Department of Psychology
George Washington University
Wasghington D.C.

ANTTI REVONSUO, PH.D.
Professor, Department of Psychology
Center for Cognitive Neuroscience
University of Turku
Turku, Finland

David M. Roane, M.D.
Associate Professor of Clinical Psychiatry
Albert Einstein College of Medicine
Chief, Division of Geriatric Psychiatry
Beth Israel Medical Center
New York, NY

R. Shayna Rosenbaum, Ph.D.
Postdoctoral Fellow
Rotman Research Institute
Baycrest Centre for Geriatric Care
Toronto, Ontairo, Canada

John R. Searle, Ph.D.
Mills Professor of Philosophy
Department of Philosophy
University of California, Berkeley
Berkeley, CA

William W. Seeley, M.D.
Clinical Fellow in Behavioral Neurology
UCSF School of Medicine
San Francisco,CA

Mark Solms, Ph.D
Chair of Neuropsychology
University of Cape Town and
Groote Schuur Hospital
Cape Town, South Africa

Donald T. Stuss Ph.D.
Director, Rotman Research Institute
Professor of Psychology and Medicine
University of Toronto
Toronto, Ontario, Canada

This page intentionally left blank

1

Introduction

TODD E. FEINBERG AND JULIAN PAUL KEENAN

Over the last decade there has been an explosion of interest in the science of consciousness. Less attention had been paid, however, to the neurobiology and neuroscience of the self. But what is "consciousness" if it is not a product of a self? It is an often ignored fact that without a self that is the subject of consciousness, consciousness does not exist; and the degree to which explicit consciousness exists depends to a large extent upon the degree that there is a subject, a self, that is the source of that consciousness.

It is not surprising, therefore, that the term *self* is as difficult to define as the term *consciousness*. According to Levin:

> The self is the ego, the subject, the I, or the me, as opposed to the object, or totality of objects—the *not me*. *Self* means "same" in Anglo-Saxon (Old English). So *self* carries with it the notion of identity, of meaning the selfsame.It is also the *I*, the personal pronoun, in Old Gothic, the ancestor of Anglo-Saxon. Thus, etymologically *self* comes from both the personal pronoun, *I*—I exist, I do this and that—and from the etymologically root meaning "the same"—it is the same I who does this, who did that. (Levin, 1992, p. 2)

The philosopher Galen Strawson, who has written about as much as any philosopher on the nature of the self, points out there is a grand multiplicity of meanings of the term *self*.

> It is difficult to know where to begin, because there are many different notions of the self. Among those I have recently come across are the cognitive self, the conceptual self, the contextualized self, the core self, the dialogic self, the ecological self, the embodied self, the emergent self, the empirical self, the existential self, the extended self, the fictional self, the full-grown self, the interpersonal self, the material self, the narrative self, the philosophical self, the physical self, the private self, the representational self, the rock bottom essential self, the semiotic self, the social self, the tranparent self, and the verbal self. (Strawson, 2000, p. 39)

That's a lot of selves! However, Strawson goes on to consider that essentially the self is, first and foremost, a *subject* of experience. To this James adds that the

self should be conceived as possessing a dual aspect as both the subject *and* an object of experience. The self as subject and object were two sides of the same coin:

> Whatever I may be thinking of, I am always at the same time more or less aware of *myself,* of my *personal existence.* At the same time it is *I* who am aware; so that the total self of me, being as it were duplex, partly known and partly knower, partly object and partly subject, must have two aspects discriminated in it, of which for shortness we may call one the *Me* and the other the *I.* (James, 1892; p. 43)

The forgoing suggest that the self is both a subject and an object of itself. In this book we consider both of these aspects of the self, but we focus on special and particular aspects of the self, namely: What happens to the self in certain neuropathological conditions? And what can these conditions teach us about the neurobiology of the self?

The next four chapters introduce the reader to some general questions regarding the philosophy and neuroscience of the self. In Chapter 2 John Searle, one of the pioneers of the philosophy of consciousness, considers first what philosophers mean by the "self." He points out that that traditionally, the problem of the self in philosophy is generally viewed as the problem of personal identity. He goes on to identify four different criteria for deciding the question of personal identity: the identity of the body, the identity of consciousness recorded in memory, the stability and continuity of personality, and "the relative coherence of the spatio-temporal continuity of the physical body through change." Searle notes that of these criteria, the problem of human consciousness poses a particular problem for our understanding of the self. He argues that considering the self as a feature of a "unified conscious field" is the best approach to understanding its ontology.

Martha Farah and Seth Gillihan then provide a selective review of the cognitive neuroscience of the self, with particular reference to imaging studies in normal subjects. The authors first discuss the difficulties encountered in performing this type of research. For example, they cast a critical eye on the definitions of the self, the interpretation of the data, the controls that are employed as comparisons for the self, as well as the methods used in analyzing neuroimaging data. The authors then divide studies of the self under four main headings: self-awareness and first-person perspective, autobiographical memory, agency, and self-concept. These distinctions allow for a succinct and cohesive review of the literature. While the authors have a number of concerns regarding current studies of the cognitive neuroscience of the self, they predict that neuroscientists will soon overcome the methodological difficulties they now encounter.

Next, Feinberg presents his model of the neurological underpinnings of the self in Chapter 4. The author has previously argued that in order to explain the unity of the self and consciousness, it should be viewed as the result of multiple, nested, hierarchically arranged neurological levels. In Chapter 4 he draws upon the prior neurological models of Maclean, Mesulum, and others and organizes the neural hierarchy of the self into roughly seven nested hierarchical levels. While the highest and most abstract aspects of the self are made possible by the hierarchically highest and most phylogenitically recent neural structures, all levels of the neural hierarchy may make a contribution to the self. Finally, he suggests that *meaning*

provides the constraint necessary for the unity of the mind's "inner eye," and *purpose* provides the constraint necessary for the highest and most intentional actions.

The next three chapters examine the self and self-related functions with reference to particular neuropsychological functions and neuroanatomical regions. In Chapter 5 Stuss and his colleagues address the frontal lobes and the self. These writers suggest, as does Feinberg in Chapter 4, that the self is hierarchically organized and propose that there are four hierarchical levels related to the self. The highest level of the self, involved in self-awareness, is subserved by the frontal lobes. These authors argue that executive processes typically associated with the frontal lobes may actually be dissociable from both self-awareness and theory of mind. They consider which candidate frontal regions might be critically involved in certain overlapping self-related functions such as autonoetic consciousness, theory of mind, and autobiographical memory, processes that they argue are more closely linked to the self than are executive functions per se.

As pointed out by Stuss and coworkers in Chapter 5, autobiographical memory is surely essential to the self as an enduring entity. In Chapter 6 Fujiwara and Markowitsch provide a further in-depth examination of autobiographical memory as it relates to the self. The authors begin by detailing the cognitive neuropsychology and neuroanatomical basis of autobiographical memory. They find that autobiographical memory is subserved by a large and interconnected network of neural structures including core memory regions such as the hippocampal formation, areas involved in self-related processing, especially the medial prefrontal cortex, and regions involved in the integration of sensory and emotional processing including the posterior association cortex and posterior cingulate gyrus. They find, in line with other authors in this volume (see, for example, Chapters 8 and 9), that nondominant hemispheric functioning plays a special role in self-related functions. They describe disorders of autobiographical memory in patients with neurological lesions as well as in patients with psychiatric disorders and consider how these clinical conditions inform us about the nature and neurobiology of autobiographical memory. In particular, they describe how autobiographical memory dysfunction in neurological and psychiatric patients converge in psychogenic amnesia and consider what this observation tells us about the relationship between autobiographical memory, emotion, and the self.

Goldenberg in Chapter 7 next considers the neuropsychological and neuroanatomical basis of the body image. He describes what he argues are the two central aspects of the body image with reference to the self: the awareness of the current configuration and permanent structure of one's own body, and the knowledge of the structure of human bodies in general. He explores the cognitive and neural bases of these two properties of the body image and along the way discusses in depth some pathologies of the body image including phantom limb phenomena, personal neglect, autoscopy, and autotopagnosia. He ultimately argues that the body image is not innate but rather acquired through experience of one's own and others' bodies.

The next six chapters focus on clinical disorders of the self. Feinberg and coworkers, in Chapter 8 describe a group of neurological disorders that bear particular relevance to the understanding of the self: *delusional misidentification* and

delusional reduplication syndromes. The authors first describe several subtypes of these conditions. Then, in an extensive review of previously published cases of these conditions, they examine their clinical, neuropsychological, and neuroanatomical features. The authors describe a particularly high incidence of right frontal pathology in these cases and consider what specific role the nondominant hemisphere might play in the creation and maintenance of the self.

In a related chapter Karen Spangenberg Postal in Chapter 9 examines cases of delusional misidentification of the self in a mirror. She describes a typical case of this disorder, examining it from the clinical, neuropsychological, and neuroanatomical points of view. She then examines the disorder in the context of psychiatry and self research as a whole. She concludes, similar to Feinberg and colleagues in Chapter 8, that the mirror sign, like other varieties of delusional misidentification, is more common in patients with right hemisphere disease and results from a complex interplay of neurocognitive and emotional factors.

In Chapter 10 Seeley and Miller consider how the self and self-related functions break down in dementia. They first present a brief overview of the phylogeny and ontogeny of the self and suggest which particular brain structures might be critical to the creation of the self. Using this model, they describe the manner in which the self may become disorganized and even dissolve in the presence of a dementing illness.

Simon Baron-Cohen next discusses the self in autism and Asperger Syndrome. Baron-Cohen, one of the leaders in this area, begins Chapter 11 with a brief introduction to autism followed by a discussion of the components of empathy. He stresses that empathy involves both cognitive and emotional aspects and describes these features of empathy within the context of Leslie's scheme of understanding minds. Baron-Cohen then considers empathy in relationship to autism and Asperger syndrome by tracing the developmental course of these disorders throughout the lifespan. The chapter concludes with a discussion of the "extreme male brain theory" previously proposed by Baron-Cohen.

Chapter 12 is an examination of schizophrenia and agency by Sarah-Jayne Blakemore. In schizophrenic patients there is a tendency to misattribute behaviors initiated by the self to an external agent. Blakemore describes experimental examinations of this tendency, focusing especially on PET imaging during voluntary action. Blakemore suggests that two regions appear important for the sense of agency, specifically the cerebellum and the parietal cortex. She suggests that excessive activity of a cerebellar–parietal network results in misattribution of agency such that self-generated actions are thought to have an external origin.

In Chapter 13 Hedy Kober and colleagues examine depersonalization as it relates to the self. The chapter addresses intriguing question regarding the brain areas that are related to disturbed self-processing in this disorder. Kober and her colleagues attempt to find common ground among studies that differ widely in method and study populations. They first examines Keenan's right-hemisphere model of the self. This is followed by an examination of the early studies of depersonalization and the brain. Modern neuroimaging studies are then considered, including experiments using PET and fMRI imaging. After a discussion of Mathew's

work (see Chapter 15) and the treatment of depersonalization disorders, the authors conclude with a description of a patient treated for depersonalization disorder using TMS.

The last series of chapters address how the self is transformed in dreams, under the influence of psychoactive agents, and during meditation. First, Antti Revonsuo examines how the sleeping brain represents the self in dreams in Chapter 14. He initially enumerates the various ways that the dream self resembles or differs from the waking self. For example, while the dream self possesses a body image like the waking self and sees the world from a similar point of view as the waking self, the dream self suffers from transient amnesia and confabulates. Revonsuo also examines the interesting feature of bizarreness in dreams and suggests that the data indicate that the dream self is in certain respects *less* bizarre than other non–self-related dream content, and he relates this observation to the patterns of REM sleep activation. In the concluding sections of the chapter, Revonsuo argues for a novel "threat simulation theory" of dreaming, in which the biological function of dreaming is a sort of "dress rehearsal" for potentially real, life-threatening events. This somewhat controversial opinion differs from the Freudian point of view that dreams often and largely serve a wish-fulfilling function.

In Chapter 15 Roy Mathew examines alterations of the self that are due to an array of psychoactive drugs. Mathew, with a decidedly non-Western emphasis, places the use of psychoactive agents into historical context and describes his own work with PET and cannabis as a model for depersonalization (also discussed in Chapter 13). After a discussion of the effects of mescaline, cocaine, and ecstasy on the self, the author makes a grand effort to relate the scientific concepts of dissociation, depersonalization, and the core self to Eastern spiritual, religious, and philosophical traditions, and all of this to the neurobiology of the self.

Hans Lou and Troels Kjaer introduce the study of meditation as a vehicle for discovering the neural correlates of the self in Chapter 16. The chapter begins with a thorough introduction to experiments designed to isolate the neural components of meditation. The authors introduce their own work on meditation in which they found precuneal, medial frontal, and striatal activation during meditation. The authors conclude that a network involving medial parietal, medial prefrontal, and right lateral parietal regions are critically involved in self-representation.

Finally, in a fascinating final chapter, world-renowned sleep and dreaming researcher J. Allan Hobson provides a harrowing yet moving personal perspective on his own "journey of the self." In February of 2001, Dr. Hobson suffered a brain stem stroke. Approximately 5 months after partial recovery from this first neurological insult, due to the combined effects of pneumonia, cardiac failure, and adverse drug reactions, Hobson went into a hallucinatory delerium. His description of this period is simultaneously fascinating and frightening and provides a marvelous window into the manner in which the mind and self can be transformed by the brain's metabolic milieu. Upon reflection Hobson concludes that in spite of his mental transformation during the time of his illness, the nature of his experiences and his ability to describe and understand them speaks to the resilience and durability of the self in the face of the ravages of neurological illness.

We hope readers enjoy the work of this eclectic group of writers and the varied approaches they take to the immensely complicated but endlessly fascinating topic of the self.

References

James, W. (1892). *Psychology: The Briefer Course.* Reprinted in: G. Allport (Ed.), *Psychology: The Briefer Course.* Notre Dame, IN: University of Notre Dame Press, 1985.

Levin, J.D. (1992). *Theories of the Self.* Washington, DC: Taylor & Francis.

Strawson, G. (2000). The phenomenology and ontology of the self. In: D. Zahavi (Ed.), *Exploring the Self. Philosophical and Psychopathological Perspectives on Self-Experience* (pp. 39–54). Amsterdam: John Benjamin.

2

The Self as a Problem in Philosophy and Neurobiology

JOHN R. SEARLE

There are a large number of different problems concerning the self in psychology, neurobiology, philosophy, and other disciplines. I have the impression that many of the problems of the self studied in neurobiology concern various forms of pathology—defects in the integrity, coherence, or functioning of the self. I will have nothing to say about these pathologies because I know next to nothing about them, and they are discussed elsewhere in this volume. I will only mention some pathologies, such as those of split brain patients, that are directly relevant to the philosophical problems of the self.

The Philosophical Problem of the Self

In philosophy, the traditional problem of the self is the problem of personal identity. Indeed, in the standard *Encyclopedia of Philosophy* (Edwards, 1967), the entry "self" just says "see personal identity." The problem of personal identity is the problem of stating the criteria by which we identify someone as the same person through changes. Thus, for example, the problem of personal identity arises in such a question as: What fact about me, here and now, makes me the same person as the person who bore my name and lived in my house 20 years ago? There are a number of criteria of personal identity, and they do not always yield the same result. I will get to these shortly.

I think that in fact there are at least two philosophical problems concerning the self. Besides the problem of personal identity, there is the problem of whether it is necessary to postulate the existence of a self that goes beyond the recognition of the body and of the sequence of experiences that occur in the body. In our philosophical tradition, and especially in our religious tradition, it is common to suppose that in addition to our bodies we also possess souls, that souls are the essence of ourselves, and that, therefore, for each of us, his or her self consists of a soul.

On this view what we think of as our mental life, both conscious and unconscious, is something that goes on not in our bodies but in our souls, which can also be called our selves or our minds. According to Descartes, an influential exponent of this tradition, each of us is identical, not with a body, but with an entity we can call mind, soul, or self, and we only happen to be contingently attached to a body during the course of a lifetime. Once we die, the soul will depart from the body and have a separate existence. I think the temptation to confuse the problem of personal identity with this second problem of the self derives from the fact that we suppose that if we had an affirmative solution to the second problem it would automatically provide a solution to the first. If we knew that in addition to our bodies we each had a soul, or self, or mind, and this entity was the very essence of our being, then the continuation of the self, so described, would immediately provide a solution of the problem of personal identity. You are identical with the person who lived here 20 years ago because you are the same soul or self.

So much for the tradition. Where are we today? Well, I do not know anybody who believes in the existence of an immortal soul except those who do so for some religious reason. A famous neurobiologist who believed in the soul was Sir John Eccles; there are a number of philosophers who also believe in the existence of an immortal soul, but like Eccles their belief is part of their general religious conviction. From my experience most philosophers do not believe in the existence of the soul. Furthermore, what is more important for the purposes of our present discussion, most philosophers do not believe in the existence of the self as something in addition to the sequence of our experiences, conscious and unconscious, and the body in which these experiences occur. I think most philosophers accept Hume's skepticism about the existence of the self (Hume, 1951, p. 251 ff). Hume asked himself the following question: When I turn my attention inward and focus on what is going on in my mind, what do I find? Hume says that I do not find any self or soul or person in addition to the sequence of my experiences. If, for example, I clutch my forehead and concentrate very seriously on what is happening in a way that will try to locate my self, what I locate will be the pressure of my hand on my forehead and a lot of other such experiences, "impressions" and "ideas" as Hume calls them. Hume's view, which has been very influential and is probably the most common view in philosophy about the self, is that each of us consists of a physical body, and each of us has a sequence of experiences within that body.[1] But that is it, as far as human life is concerned. There is no self or soul left over, nor is there any need to postulate any such entity.

Well, what about personal identity? There are a variety of criteria that we do, in fact, employ in deciding questions of the identity of a person across time and change. It seems to me that, in fact, we employ at least four different criteria for deciding questions of identity. The first and most important is the identity of the body. I am the same person as the person who bore my name decades ago because my present body is spatiotemporally continuous with the body that existed under my name at that time. Of course, there are philosophical puzzles: None of the molecules in my body today is the same as those in my body of decades ago, so how can the body be the same if all the microparts are different? Furthermore, philoso-

phers are good at inventing puzzling science fiction thought experiments. Suppose that bodily fusion and fission were common. What would we say if humans routinely split into two or three or five bodies, as amoebae now split into two? But in spite of these puzzles, we have a pretty clear notion of bodily identity that works across time and change. Well, why isn't that enough? Unlike the identity of material objects such as cars and houses, we are convinced that the identity of the body is not enough to constitute a personal identity. We all understand Kafka's story of Gregor Samsa who wakes to find *himself* in the body of a giant insect, and it is easy to imagine science fiction scenarios of brain transplants in which I might find myself having a different body. Furthermore, possession of the same brain might not by itself be enough for personal identity. Suppose that I had the same brain but that all the information in my brain were transferred to another person's brain, and the information in his brain were transferred to mine. We might feel that I now inhabit his body, and he inhabits mine. I am not saying that these science fiction fantasies are sufficiently clear, or even coherent. I only point to them because they indicate that where our own personal identity is concerned, we think there is more to it than just the body.

Well, what more? Locke said that the essential thing to personal identity is what he called "consciousness" (Locke, 1947, pp. 182–201). Most interpreters think that by *consciousness* he meant our present memory experiences of a continuity between our present self and the earlier self that had the experiences on which our present memories are based. In short, Locke's consciousness criterion is usually, and I think correctly, interpreted as a memory criterion. The idea is that in addition to the continuity of the body, we need a continuity of consciousness as recorded in memory. In addition to the third-person criterion of bodily continuity, we need the first-person criterion of the experience of the personal identity of the self. And this is how all human personal identity differs from the identities of cars, houses, etc.

A third criterion, commonly used in ordinary life, is relative stability and continuity of personality. In cases in which we feel that a person's personality has altered dramatically and drastically, we are inclined to feel "she is not the same person any more." To take a famous case, when an iron bar went through the skull of the nineteenth century railway worker Phineas Gage, he miraculously survived, but his personality was totally different. Before he had been friendly, gregarious, and reliable; afterwards he became hostile, surly, and capricious. From a purely practical point of view, we would continue to regard him as the same person. For example, he would still owe the taxes of Phineas Gage and still own the property of Phineas Gage, but from a neurobiological point of view and a philosophical point of view, we would want to know very much what had changed in Phineas Gage so as to render him a totally different personality from what he had been before.

A fourth criterion is the relative coherence of the spatiotemporal continuity of the physical body through change. There is a standard pattern by which one and the same body grows and ages until eventual death, but suppose that the entity, though spatiotemporally continuous, varies wildly and unpredictably in its physical form. Suppose my body might change into that of a car or a house or a mountain. We think we understand Gregor Samsa's body changing into that of a large insect, but how

far are we prepared to go? I do not think we need to answer that question in advance. The point I am making now is that we in fact employ four different sets of criteria in our concept of personal identity—spatiotemporal continuity of body, continuous memory, continuity of personality, and coherence of physical change—and that the everyday concept works well enough because these hang together to give consistent answers in real life.

So far so good, or so it might seem. It seems there is no such thing as the self in addition to all the stuff I have been talking about—continuity and coherence of the living body together with continuous memory sequences and coherent personalities, but I do not think this conclusion is correct. I have reluctantly come to the conclusion that the nature of human consciousness requires the postulation of a non-Humean self, and this postulation poses problems for neurobiology that go beyond the standard neurobiological problems of consciousness but will enable us to re-pose the question of consciousness in important ways.

The Neurobiological Problem of Consciousness

Sometimes, but unfortunately not very often, we can get a scientific solution to a long-standing philosophical problem. A famous case is the problem of life. The problem was: how can mere inert, inanimate matter be alive? Traditionally, there were two possible answers, the mechanist answer, according to which life could be reduced to mechanical processes, and the vitalist answer, according to which something more was needed, an *élan vital,* a vital force, that infused life into inert matter. We cannot take this problem seriously anymore, and it is hard for us to recover the passion with which it was debated a mere century ago. The point is not that the mechanists won and the vitalists lost, but rather that we got a much richer conception of biochemical mechanisms—a conception that did not exist when the debate raged in the nineteenth and early twentieth centuries.

I hope something like this is also happening to the problem of consciousness. The problem here is: how can mere unconscious bits of matter in the brain cause consciousness? On this problem we have a head start over the problem of life because we know before we ever get started on the investigation that processes in the brain do, in fact, cause consciousness. All the same, much, though not all, current neurobiological research suffers from a mistaken conception of the problem, and that in turn derives from a mistaken conception of the self. In order to work up to the self, I have to say a bit about consciousness.

I sometimes still hear it said that "consciousness" is hard to define. But if we are just talking about a definition that gives us not a scientific analysis, but rather locates the target of our investigation, then it seems to me that consciousness is not hard to define. Here is a definition: consciousness consists of those states of feeling, sentience, or awareness that typically begin when we wake from a dreamless sleep and continue throughout the day until those feelings stop, that is, until we go to sleep again, go into a coma, die, or otherwise become "unconscious." On this account dreams are a form of consciousness that occur to us during sleep. What,

then, are the features of consciousness that we would like to be able to explain on this definition? Conscious states, so defined, are *qualitative* in the sense that there is always a certain qualitative feel to what it is like to be in one conscious state rather than another. We all know the difference between listening to Beethoven's Ninth Symphony and drinking cold beer. The difference is precisely the kind of qualitative difference that I am talking about. We know furthermore that all such conscious states are *subjective* in the sense that they exist only as experienced by a human or animal subject. Conscious states require a subject for their very existence. They do not exist in a neutral or third-person fashion; they have an existence that depends on their first-person subjective qualities, and that is just another way of saying that a conscious state must always be someone's conscious state. In philosophy this point is sometimes put by saying consciousness has a "first-person ontology." *First-person* here means there must be an *I,* some subject that experiences the consciousness, and *ontology* just refers to the mode of existence that something has. A third feature of consciousness is less frequently remarked on, but I think it is absolutely essential to understanding the other two. Conscious states always come to us as part of a unified conscious field, so when I am listening to Beethoven's Ninth Symphony while drinking beer, I do not just have the experience of listening and the experience of drinking; rather, I have the experience of drinking and listening as part of one total conscious experience, and this is characteristic of consciousness generally, that consciousness always and only occurs as part of a unified conscious field. This is why, by the way, the split brain experiments are so important to the study of consciousness. As far as we can tell from the experiments of Sperry and Gazzaniga (Gazzaniga, 1985), a patient whose corpus collosum has been cut gives all the external symptoms of having two separate conscious fields, one in each hemisphere, and these are only imperfectly united into a single conscious field; sometimes they exist as separate conscious fields.

Among philosophers, Immanuel Kant attached a great deal of importance to the unity of the conscious field. He called it "the transcendental unity of apperception" (Kant, 1997). I think the unity of our conscious field is important to our analysis of the concept of the self, and I will say more about it later. For the moment, I just want to call attention to the fact that these three features, qualitativeness, subjectivity, and unity, are not independent of each other. Each implies the next. You cannot have a qualitative experience such as tasting beer without that experience occurring as part of some subjective state of awareness, and you cannot have a subjective state of awareness except as part of a total field of awareness, even if the only thing in this particular impoverished field is the state of awareness itself. So we might say, initially at least, that the problem of consciousness is precisely the problem of qualitative, unified subjectivity. The three features are simply different aspects of the one common essential trait of consciousness. Now there are lots of other traits of consciousness that should be investigated, and I have investigated the philosophical aspects of them at some length in a number of books (Searle, 1984, 1992, 1997, 2004). However, for the purposes of this chapter, I will focus on only these three, and particularly on the last, because they are most relevant for our examination of the problem of the self.

Notice an interesting feature of the unified conscious field. Within the field we can change our attention at will. Without moving my head or even my eyes, I can focus my attention on this or that feature of my visual field. And even with my eyes closed I can think now about this problem, now that problem, moving the focus of my attention, again, entirely at will. This ought to seem puzzling to us. The brain creates a conscious field just as the stomach and digestive tract create digestion. So what has conscious *will* got to do with it? To put the question crudely, when I say I can shift my attention at will, who does the shifting? Why should there be anything more to my conscious life than the existence of a conscious field? Where is there anything more? I will come back to these questions because I think they are essential to understanding the problem of a self.

How can we solve the problem of consciousness as a problem in neurobiology? First of all we have to state exactly what the problem we wish to be able to solve is, and here I think the answer can be stated quite simply. The neurobiological problem of consciousness is: How exactly do brain processes cause our conscious states in all their enormous richness and variety, and how exactly are these conscious states realized in the brain? Why do conscious states exist at all, and where and how do they exist in the brain? It took a long time for many neurobiologists to see that this was a crucial question in neurobiology; indeed, I would say it is the number one question in the biological sciences today. Right now there is a great deal of research on precisely this topic.

Most researchers are seeking the neuronal correlate of consciousness (NCC). The idea is this: in order to solve the problem of consciousness, we should find out first what is going on in the brain at the neurobiological level at a time when a subject is conscious. What neurobiological features are correlated with the conscious features? We now think, perhaps with too much optimism, that recent improvements in our investigative techniques, especially single-cell recording and fMRI, will give us a richer research apparatus for discovering the NCC. The idea, though often not explicitly stated, is that the investigation will proceed according to a pattern that has been fairly common in the history of science. The first step is to find a neuronal correlate of conscious states. This would be the NCC. The second stage is to investigate whether the NCC is actually a *causal* correlation, and we do this by the usual tests. In an otherwise unconscious subject, can you produce consciousness by producing the NCC? In an otherwise conscious subject, can you shut off consciousness by shutting off the NCC? If you have affirmative answers to these questions, then it is a reasonable supposition that the correlation is more than accidental; it is, in all likelihood, a causal correlation.

The third stage, and we are a long way from reaching it, is to formulate a general theory, a general statement of the laws or principles by which the correlation functions causally in the life of the organism. This research, as I said, is off and running. I am quite optimistic about its long-term prospects, though I have to admit progress has been very slow. In general, there are two lines of research that go on in this field, one of which seems to me much more promising than the other, although the more promising, unfortunately, is harder to conduct as an actual research project. The most common line of research is what I call the "building-block

approach" (Searle, 2000). The idea of this approach is to think of the unified conscious field as made up of all of its different components. Right now, for example, I am experiencing the color red as I look at a red box on my table, I am hearing the sound of my voice, I am feeling a slight aftertaste of coffee in my mouth, and so on. The idea of the building-block approach is to think of the entire conscious field as made up of such building blocks (the experiences of color, of sound, of taste, etc.). On this view, if we could find the NCC for even one building block and understand the mechanisms by which that NCC caused consciousness, that presumably would give us an entering wedge that would enable us to crack the whole problem of consciousness. The mechanisms by which the NCC for a particular conscious state produce that conscious state will presumably be generalizeable to other conscious states. The analogy with genetics is obvious: You do not have to know how every phenotypical trait is the expression of some gene or set of genes in order to appreciate the power of the DNA conception of genetics. You have to understand the general mechanisms involved, and then you can apply them to particular cases. Most research on consciousness that I am aware of follows the building-block approach.

Another approach, pursued by a minority of investigators, is what I call the "unified-field approach." We want to know not so much what causes the experience of red, though that is part of our overall investigation, but rather how the brain becomes conscious in the first place. What exactly is the difference between the unconscious brain and the conscious brain, and how exactly do those differences cause the brain to be in a state of consciousness? The state of consciousness, as I have argued earlier, is a matter of a unified conscious field, so the question for this approach is: how does the brain produce the unified conscious field?

I said that I think the unified field approach is superior. Why? Science typically has proceeded by the practice of breaking larger problems down into smaller problems by using an atomistic approach to large problems. Why would this not work for consciousness? Perhaps it will, but there is an immediate objection: the building-block approach identifies building blocks that can exist only in a subject who is already conscious, but if that is right then it looks as if the NCC for the experience of the color red does not give us the NCC for the experience of consciousness; rather, it gives us the NCC for a particular mode *within a preexisting conscious field.* On the unified field approach we should think of perception *not as creating* consciousness, but as *modifying* the preexisting conscious field (Llinas, 2001). On the building-block approach perception creates consciousness just like that, out of nothing except neuronal processes. On the unified field approach perception does not create consciousness but modifies the consciousness of the preexisting conscious field.

Why am I so convinced that the building-block approach is the wrong approach? The answer is that if we take the building-block approach as giving us the NCC for consciousness and not for particular modifications of the conscious field, then it would make predictions that seem implausible. The approach would predict that in an otherwise unconscious subject, if you could introduce the NCC for the experience of the color red the subject would suddenly have a conscious flash of red and then lapse back into total unconsciousness. That seems to me extremely un-

likely. From what we know about the experience of red, it occurs only in subjects who have a preexisting consciousness, that is, who are conscious already when they experience red, and so on with perception in general. Alarm clocks, for example, do not create just a single percept but rather create a field in which that percept is the central entity.

Whether the building-block or the unified-field approach is better is an empirical question not to be settled by philosophical analysis, and I am prepared to be proven wrong. Perhaps the building-block approach will succeed in the end, but right now I think it is a source of difficultly. In fact, it turns out it is not at all hard to find various kind of NCCs for particular sorts of experiences, and many researchers have done that (Kinwasher, 2001). But we still have not solved the problem of consciousness by these findings because we still do not have an answer to the question: what makes the brain conscious? The reason I have belabored this point is because I think there are lessons to be learned about the neurobiological problem of the self from reflecting on the neurobiological problem of consciousness.

The Requirement of the Self as a Formal Feature of the Unified Conscious Field and Its Implications for Neurobiology

There are famous objections to Locke's idea of memory as the essential criterion for personal identity. One objection is this: It would be circular to make memory a criterion for the identity of the self, because in order to establish that the memories in question are correct memories, one first has to establish that the person who has these memories is really identical with the person whose experiences he claims to remember. Thus, if I now sincerely claim that I remember writing the *Critique of Pure Reason,* that by itself goes no way at all toward showing that I am, in fact, identical with the actual author of the *Critique of Pure Reason,* because one would first have to establish that I did write the *Critique* before one could know that the memories are accurate. For exactly the same reason, the fact that I now claim to remember writing *Speech Acts* by itself goes no way at all toward showing that I am identical with the actual author of *Speech Acts.* Hence, it looks as if memory is no good as a criterion of the self because, to establish that the memory is an accurate memory as opposed to an illusory one, one first has to establish the very identity that the memory was supposed to establish. I think this is a fair objection if we treat memory as a criterion of personal identity, but that need not be our only interest in memory. It seems to me, for this discussion, that what we are interested in is not how to establish conclusively that I am identical with such and such a person who lived so many years ago, but rather what facts about my conscious states give me a sense of myself as a single continuing entity through time? It is this sense of the self that is more relevant to problems in neurobiology.

I now think that with the introduction of memory I am prepared to state the philosophical problem of the self, and how it bears on neurobiology, a little more precisely. It is a remarkable feature of the conscious field, which I identified earlier, that the elements of the conscious field are not, so to speak, neutral. They are

not just given to me as independent phenomena, but rather they exhibit certain special traits that I now will to specify further. First, it is an absolutely astounding thing about the conscious field that, given the same conscious field, I can shift my attention at will. Even without changing the direction of my eyes, I can focus my attention now on the coffee cup on the table, now at the computer screen in front of me, now at the bookcase on my right. The shift of attention within a constant conscious field is something I can do at will. A second feature, which derives from the first, is that I can change the entire conscious field at will, simply by doing something different, such as moving my head, or closing my eyes, or standing up and leaving the room. The fact that I have the ability to do things seems to be an essential part of the normal human conscious field, and we can easily imagine a different mode of existence in which I was utterly passive and I simply experienced events occurring to me but had no sense whatsoever of having any control over them. When I engage in conscious voluntary action, I have a sense of my own freedom. I have the sense that I am doing this, but I could, right here and now, be doing something different. In such cases I have the impression that the causes of my action, in the form of the reasons on which I am acting, are not causally sufficient to determine the action. In normal nonpathological cases the action is *motivated but not determined,* because there is a *gap* between the perceived causes and the action. This gap has a name in philosophy, it is called the freedom of the will. It does not matter for our present purposes whether the sense of freedom is a mark of real freedom or only an illusion. I cannot think the gap away, for even if I become a convinced determinist and refuse to make any choices on the grounds that everything is determined anyway, my refusal to make any choices is intelligible to me as my action only under the presupposition of freedom. I have freely chosen not to make any free choices. The third feature of the conscious field is that I do, in fact, have a sense of myself as a particular person situated at a particular time and place in history with a certain set of particular experiences and memories. We need to put these various features together into a unified account of the self before we can state questions that could be addressed by neurobiology.

The sequence of conscious experiences (as identified by Hume) together with the fact that these experiences come to us as part of a unified conscious field (as identified by Kant) is still not enough to give us the characteristic experiences that constitute our idea of the self. Even if we add to the Hume-Kant story the idea that some of these experiences are memories of earlier experiences (as identified by Locke), we still do not have our conception of the self. What is missing? Let us go back to the point I made earlier, that we can shift our attention at will and indeed initiate actions at will. Who does the shifting, and who does the initiating? One thing I have noticed in teaching these matters to undergraduates and discussing them with professionals is that *everybody* feels the attraction of the homunculus fallacy. It is very tempting to think that there is a little guy in my head who does my thinking, perceiving, and acting. Of course, the homunculus fallacy is a fallacy, because it leads to an infinite regress. If my vision can only occur because the little man in my head watches the TV screen in my head, then who watches the TV screen in the little man's head? But, and this is the crucial point, though the ho-

munculus is a fallacy, the urge to postulate the homunculus is powerful and well founded. The problem is that we cannot make sense of our conscious experiences if we think of them as just a sequence of events (impressions and ideas à la Hume) related by present memory experiences of earlier experiences (à la Locke) and part of a unified conscious field (à la Kant). We need to postulate, initially at least, a locus of the initiation of action. My decisions and actions are not just events that occur, but rather *I decide* and *I act.* But now we have to proceed very carefully, or else we will start sounding like the worst kind of German philosophers (Was ist das Ich?). So far we have postulated only a purely formal entity. It is simply an x, something capable of initiating and carrying out actions. Notice, however, that the entity that initiates actions must be the very same entity as the entity that reflects on reasons for action, and indeed the same entity that has perceptions and memories that form the basis of the reasons on which it reflects and decides on actions. Just as we had to postulate a purely formally specified entity that decides and acts, so the connection between perception, memory, and reasons for action requires us to postulate that the *same* entity that performs the action has all of these other features. Why? Well, if the entity that decides and acts is different from the one that perceives, remembers, and reflects, then we would not get the connection necessary to make sense of our actions. If I act on a reason R, then R must be *my* reason for acting. For example, if I jump out of the way because I see a truck bearing down on me, then the entity that initiates the jumping has to be the same one that does the seeing, otherwise the seeing gives no reason for the jumping. Furthermore, once the action has been performed, the same entity that did the performing is the one who has responsibility for the performance and thus gets the credit or the blame. We can pull all these threads together as follows.

The universal urge to postulate a homunculus is based on very profound features of our ordinary conscious experiences. In order to make sense of those experiences we have to suppose,

There is some x such that

x is conscious

x persists through time

x has perceptions and memories

x operates with reasons in the gap

x, in the gap, is capable of deciding and acting

x is responsible for at least some of its behavior.

The x in question is the self in at least one sense of the word. Notice that the postulation of the self is not the postulation of a separate entity distinct from the conscious field but rather it is a formal feature of the conscious field. The point I am making is that if we reflect on the features of the conscious field, we see that we cannot accurately describe it if we think of it as a field constituted only by its con-