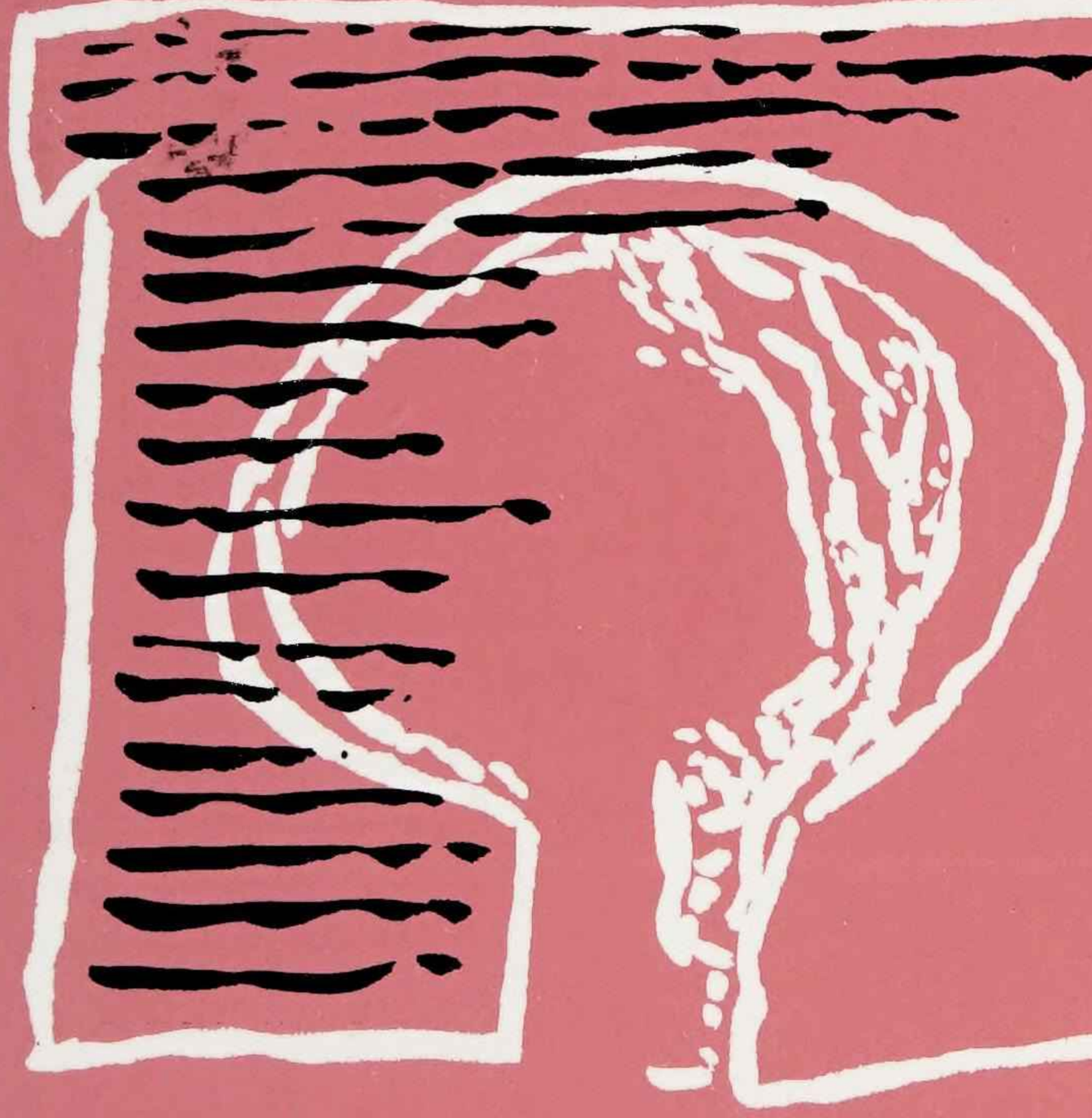
ADVANCES IN CONSCIOUSNESS RESEARCH



Consciousness and qualia

Leopold Stubenberg

Digitized by the Internet Archive
in 2019 with funding from
Kahle/Austin Foundation

https://archive.org/details/consciousnessqua0000stub

CONSCIOUSNESS AND QUALIA

ADVANCES IN CONSCIOUSNESS RESEARCH

ADVANCES IN CONSCIOUSNESS RESEARCH provides a forum for scholars from different scientific disciplines and fields of knowledge who study consciousness in its multifaceted aspects. Thus the Series will include (but not be limited to) the various areas of cognitive science, including cognitive psychology, linguistics, brain science and philosophy. The orientation of the Series is toward developing new interdisciplinary and integrative approaches for the investigation, description and theory of consciousness, as well as the practical consequences of this research for the individual and society.

EDITORS

Maxim I. Stamenov
(*Bulgarian Academy of Sciences*)

Gordon G. Globus
(*University of California at Irvine*)

EDITORIAL BOARD

David Chalmers (*University of California at Santa Cruz*)
Ray Jackendoff (*Brandeis University*)
Christof Koch (*California Institute of Technology*)
Stephen Kosslyn (*Harvard University*)
Earl Mac Cormac (*Duke University*)
George Mandler (*University of California at San Diego*)
John R. Searle (*University of California at Berkeley*);
Petra Stoerig (*Universität Düsseldorf*)
Francisco Varela (*C.R.E.A., Ecole Polytechnique, Paris*)

Volume 5

Leopold Stubenberg

Consciousness and Qualia

CONSCIOUSNESS AND QUALIA

LEOPOLD STUBENBERG
University of Notre Dame, Indiana

Thomas J. Bata Library
TRENT UNIVERSITY
PETERBOROUGH, ONTARIO

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

BD418.3 .S78 1998

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences — Permanence of Paper for Printed Library Materials, ANSI Z39.48–1984.

Library of Congress Cataloging-in-Publication Data

Stubenberg, Leopold.
Consciousness and qualia : Leopold Stubenberg.
p. cm. -- (Advances in consciousness research, ISSN 1381-589X ; v. 5)
Based on the author's thesis (doctoral)--1992.
Includes bibliographical references and index.
1. Qualia. 2. Consciousness. I. Title. II. Series.
BD418.3.S78 1998
126--dc21 98-14824
ISBN 90 272 5125 8 (Eur.) / 1-55619-185-5 (US) (Pb; alk. paper) CIP

© Copyright 1998 - John Benjamins B.V.
No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. • P.O.Box 75577 • 1070 AN Amsterdam • The Netherlands
John Benjamins North America • P.O.Box 27519 • Philadelphia PA 19118-0519 • USA

The form of a philosophical theory, often enough, is:
Let's try looking over here.

Jerry Fodor

Table of Contents

1. Introduction 1

2. Making our Ideas Clear: Consciousness and Qualia 12

3. On Method: The First-Person Perspective 36

4. Higher-Order Representation and Introspection 59

5. The Allure of Introspectionism 93

6. Oscar, the Unconscious Introspector: A Case Study 129

7. Relocating Qualia 151

8. Having Relocated Qualia 181

9. Denying Relocated Qualia 214

10. Consciousness: The Having of Qualia 262

Notes 311

References 332

Detailed Table of Contents 346

Index 352

List of Tables and Figures

Table 1. Four ways of deMenting qualia 153
Table 2. Preliminary chart of representation-based accounts 184
Table 3. Final chart of representation-based accounts 184
Figure 1. Dretske's spot 208
Figure 2. The Muller-Lyer Illusion 252
Figure 3. The Benham Disk 255

Acknowledgments

Over a period of many years a large number of people have given me the various kinds of hlep I needed to write this book. Among them are teachers, mentors, colleagues, and friends. Many of them have played more than one role. To all of them I am very grateful. They are Bill Lycan, Bill Ramsey, Chris Maloney, Dave Chalmers, David Armstrong, Dean Zimmerman, Fritz Warfield, Gordon Globus, Hannes Brandl, Hannes Marek, Heiner Rutte, Jennifer Lackey, John Pollock, Joseph Tolliver, Keith Lehrer, Ken Sayre, Maxim Stamenov, Rob Cummins, Robert Rogers, Rudolf Haller, Scott Sturgeon, and Trenton Merricks. I am especially grateful to Brigitte Prutti and Marian David. The book and its author would be in much worse shape without them.

This book is based on a Ph.D. dissertation of the same title completed in 1992. Section 3.3.1 includes material first published in Stubenberg (1993). Section 3.3.2 is based on a talk contributed to a conference on Daniel Dennett's philosophy of mind, Notre Dame Spring 95. Chapter Six is a revised version of Stubenberg (1992b). Some of the ideas in Section 10.6 appear in Stubenberg (1996) and Stubenberg (1997).

CHAPTER ONE

Introduction

1.1 Philosophy or Science

"Why, if it weren't for this 'internal illumination'...the world would be nothing but a pile of dirt!"[1] The meaning of Einstein's sage words is easily felt but hard to explain. I will try to explain what the internal illumination is. Why its absence should turn the world into something so undesirable I do not attempt to answer.

What is it to explain the internal illumination? One might, for example, attempt to determine the neurophysiological conditions and mechanisms that give rise to the internal illumination. Only a trained neurophysiologist could undertake this project. As a 'pure' philosopher I am in no position to carry out such a task. But there is something else the philosopher can do. Before anybody can reasonably attempt to present a scientific explanation of the internal illumination somebody has to tell us what sort of phenomenon the internal illumination is supposed to be. That such an account of the internal illumination is needed is easily overlooked. The light is, after all, switched on in everyone of us. So it is only natural to think that we know, in a particularly intimate way, what the nature of the internal illumination is. But experiencing a phenomenon is not the same as understanding it. In the case of the internal illumination, we get the former for free while the latter is surprisingly difficult to attain. In this book I attempt to advance our understanding of the internal illumination by presenting an analysis of it.

A comparison with the phenomenon of knowledge may be useful. For knowledge is better understood than the internal illumination. We all know many things but few of us are in a position to present an analysis of knowledge. A scientific study of knowledge must therefore be prefaced by a philosophical analysis of the concept of knowledge. For if one does not know what knowledge is one cannot look for it. This point holds quite generally. If, for example, you do not know what toadstools are you cannot go and look for them. The classical philosophical analysis of knowledge has it that knowledge is justified, true be-

lief. Each of these three terms requires further analysis. Let us focus on justification. All the points made above about knowledge apply to justification as well. We all have many justified beliefs. Yet, few of us are ready to present an analysis of justification. But without analysis of justification we are in no position to study the phenomenon scientifically. One of the many competing analyses of justification construes it in terms of reliability: a belief is justified just in case it was produced by a reliable mechanism. With this step of the analysis, we approach a point where philosophical analysis and scientific investigation meet. For it seems obvious how the scientific study of the human cognitive architecture bears on the question whether humans have reliable belief producing mechanisms.[2]

As yet we find no similarly pleasing convergence of the philosophical and the scientific study of the inner light. The philosophers are to blame. In recent years competing analyses of the inner light have multiplied in an unprecedented manner. Using a very broad brush stroke, the present situation presents itself like this: Of the many current analyses some accurately capture the inner light and some do not; and some offer practicable suggestions for scientific research and some do not. Those analyses that are accurate have no scientific implications and those that readily suggest scientific research projects are not accurate. My own proposed analysis strives to accurately capture the phenomenology of the inner light. But it does not attain the level of concreteness that might usefully suggest a direction for scientific research.

1.2 Outline One: Some Key Ideas

The problems I want to address are not new. They first arose more than two thousand years ago when Democritus said that "by convention is color, and by convention sweet, and by convention every combination, but in reality the void and atoms." (Barnes 1982: 371) This set into motion the long retreat of perceived, sensory qualities from the world. Temporarily, these qualities found a resting place in the Mind.[3] On John Locke's version of this doctrine secondary qualities exist only as ideas:

> Since were there no fit Organs to receive the impressions Fire makes on the Sight and Touch; nor a Mind joined to those Organs to receive the *Ideas* of Light and Heat, by those impressions from Fire, or the Sun, there would yet be no more Light, or Heat in the World, than there would be Pain if there were no sensible Creature to feel it, though the Sun should continue just as it is now, and Mount *Ætna* flame higher than ever it did. (Locke 1982: II, XXXI, 2)

By removing these troublesome properties from the world and locking them up in the Mind the world is opened up for scientific, naturalistic explanation. But this solution is only temporary. For the scientific picture of the world is incomplete as long as it does not encompass its creators. The drive towards a scientific, naturalistically respectable theory of ourselves marginalizes the Mind. Unlike the sensory qualities, the Mind cannot be saved by being swept into the Mind. The endpoint of this development is the elimination of the Mind and everything that it contains. Thus the sensory qualities that first were banished from the world have now vanished altogether. Only the most unafraid of philosophers have dared to embrace this radical conclusion. One recent proponent of this view (as it applies to color) has argued that visual perception involves a "collective hallucination." (Landesman 1989: 114) Our error is not merely a matter of mistakenly projecting the inner colors onto the outer objects—"for there are no exemplified qualities to project." (Landesman 1989: 109) The hallucination consists in the fact that "all visual perception presents us with colors and colored objects, although in fact nothing at all exemplifies color." (Landesman 1989: 115)

Consideration of this radical eliminativism helps to bring out in full relief a basic assumption underlying this investigation as well as one of the two central questions I try to answer. Bluntly stated, the assumption is this: eliminativism is unacceptable. Bruce Aune has expressed the intuition underlying this assumption in a way that I could not possibly improve upon:

> I am currently looking at the reddish cover of a book, and I have no doubt about the existence of a reddish expanse in some sense "before my mind." In fact, I have never had better evidence for the existence of anything than I now have for the existence of the expanse: I have *vastly better evidence for it* than I have for the physicist's current belief about the ultimate constituents of matter. (Aune 1989: 12-13)

There are questions about how to best conceptualize this datum. But analyses that make the datum disappear are ruled out. This basic assumption must not, however, blind us to the fact that the problem to which eliminativism is a reaction is genuine and deep. Once the Mind is out of the running, nothing in the world seems fit to exemplify the qualities with which experience acquaints us. If eliminativism is to be avoided, we must find a place for sensory qualities in the world. Put another way, we need to determine the nature of sensory qualities—What are they and what do they qualify? I call this the 'nature-problem.'

In order to introduce the second big question, we need to backtrack a little. There are those who find fault with my mini-history of sensory quality from Democritus to Landesman. They hold that sensory qualities in general and colors in particular are perfectly respectable properties of physical objects. Take, for

example, the reddishness before Aune's mind. There is some discussion about exactly what kind of property of what kind of object Aune's reddishness is. But the main point is this: The red color that suffuses Aune's visual field when he looks at the book is a physical property in good standing. There is nothing strange or nonphysical about it. No mind is needed to contain it. According to this view my 'big philosophical problem' of finding a place for sensory qualities in the world is simply one of the many factual questions that color science will eventually answer.

There is something reassuringly levelheaded about this view. For the sake of the argument, I grant the assumption that the well-known problems confronting the identification of sensory qualities with physical properties can be overcome. But there are other problems. Consider, again, Bruce Aune looking at the book in front of him. Assume that the red color he sees is a complex reflectance property exemplified by the book's cover. What has to transpire between the book and Aune so that Aune ends up having the characteristic sort of experience that we call seeing red? The simplest answer is this: Aune must represent the book's color in order to conjure up the reddish expanse before his mind. An answer that I am particularly interested in pursuing is a bit more complex: Aune must represent the book's color, and he must represent his representation of the book's color. Only with the second-order representation (i.e., with the advent of introspection) does the red expanse unfold itself before his mind. Now, imagine that next to Aune there is a video camera mounted on a tripod. The film is rolling and it, too, is representing the book's color. Feel free to integrate a second, smaller camera into the first camera. It represents the representation produced by the first camera. I think that nobody will be inclined to judge that the camera (with or without the introspective module) experiences the book in the way in which Aune presumably does. Using a notorious bit of philosophical jargon, we can describe the difference between Aune and the camera as follows. For Aune, but not for the camera, it is like something to look at the red book. And this salient difference between Aune and the camera exists even though their representational structures are analogous. Hence the fact that there is a red expanse before Aune's mind (or alternatively, the fact that Aune has a red experience or the fact that it is like something for Aune to look at the red book) does not consist in his representing, or metarepresenting, the red book in front of him. For the camera does that too. Yet it is like nothing at all to be a camera.

The point of this comparison of Aune and the camera is the following. To say that the color red is a feature of the perceived object is only the first step in an explanation of the sensory redness we experience. In the second step one has to explain how the perceiver manages to get at the book's redness in the right sort of way—in a way, that is, that explains why bearing this sort of relation to the

red book should result in there being a red expanse before his mind. The little story about Aune and the camera suggests that it will be difficult to rise to the challenge posed by the second step. I call this the 'having-problem.'

We now have before us, in rough outline, the two questions I want to answer. What sort of property is the redness of Aune's red expanse? And, How does it get to be before his mind? Abstracting away from the particulars of Aune's case and using the terminology employed throughout the rest of this book, I can restate these questions as follows. What are qualia? How are qualia had? An answer to the first question is a solution to the nature-problem. An answer to the second question is a solution to the having-problem. Taken together these answers amount to an account of qualitative consciousness. For to be conscious just is to have qualia. In other words, the fact that there is something it is like to be you consists in the fact that you have qualia.

The thesis I just stated—to be conscious is to have qualia—is the thesis I want to defend. A few words about the nature of this thesis and the manner of its defense are in order. The thesis states a result of philosophical analysis. As it stands it is virtually empty. For we do not yet know what qualia are and what it is to have them. Defending the thesis is a matter of filling out this hollow schema by providing substantial accounts of qualia and the having-relation. A survey of a number of extant theories of consciousness shows that such accounts are not easy to come by. The task is made more difficult by the self-imposed constraint that these accounts should fit within a broadly naturalistic framework.

This approach suggests that we begin by asking two interconnected questions of any purported theory (T) of consciousness. First, Do the qualia of T fit into a naturalistically conceived world? Second, Does the having-relation of T fit into a naturalistically conceived world? If the answers to these questions are Yes and Yes, we can go on to ask the crucial third question: Does the fact that a person has (in this manner) a quale (of this sort) make it clear why being so related to such a thing results in it's being like something to be that person? If the answer to this question is Yes, too, we have a successful theory of qualitative consciousness.

Summarizing my conclusions in telegram form I can say this. All the theories I criticize display the following pattern of answers to the above three questions: Yes!, Yes?, and No. They all make qualia into respectable, broadly physical properties. They all rely, in one form or another, on the machinery of representation to explain the having-relation. This raises the large question of the naturalizability of representation. I do not, however, discuss this question here and simply go along with the assumption that representation can be naturalized. But affirmative answers to the first two questions only guarantee that the account is naturalistically acceptable. This is a necessary but not a sufficient condition for

success. The crucial third question concerns the issue of phenomenological adequacy. And on this count all of the criticized theories fail. These theories do not make it clear why it should feel like anything to represent the very plain qualia they posit.

The problem underlying this pattern of failure is this: Qualia, as traditionally conceived, have a very bad reputation among naturalistically inclined philosophers. Much concerned to avoid properties whose nature is thoroughly mysterious they defrock qualia of all their traditional privileges and make them into lowly properties of material structures. This solves the nature problem. But once the qualia have sunk so deep, no nonmysterious 'having' mechanism is powerful enough to raise them back up sufficiently to make qualitative consciousness out of them.

This analysis suggests a strategy of integration. I want to resist reductionism about qualia by arguing that the naturalist's world is spacious enough to accommodate these peculiar properties in unreduced form. What I need to show then is that "the red glow of the sunset [is] as much part of nature as are the molecules and electric waves by which men of science would explain the phenomenon." (Whitehead 1926: 29) One cannot help but agree with Whitehead when he speaks of the "extreme difficulty of exhibiting the perceived redness and warmth of the fire in one system of relations with the agitated molecules of carbon and oxygen, with the radiant energy from them, and with the various functionings of the material body." (Whitehead 1926: 32) The proposed integration of qualia into the physical world resists summarization that is both brief and intelligible. But as regards brevity, Russell's provocative claim that "the brain consists of thoughts" (Russell 1975: 18) cannot be outdone.[4] Anticipating the reader's reaction, Russell adds: "To this people will reply 'Nonsense! I can see a brain though a microscope, and I can see that it does not consist of thoughts but of matter just as tables and chairs do.'" (Russell 1975: 19) I hope to show that this objection is not as powerful as it seems. The reason is that what you see when you 'see a brain' (or a table or a chair or any material object, for that matter) is not at all what you think you see. But the details of the proposed solution to the nature-problem will have to await the last chapter.

Let us assume, for the sake of argument, that the red expanse (which Aune takes to be before his mind) is, somehow, in Aune's brain. Does its being there explain why this arrangement should result in there being something it is like to be Aune? I think not. We still need an account of how the red expanse is being *had*—how it gets to be before Aune's mind. Having lost all faith in representation-based accounts of the having-relation, I need to look somewhere else. The crucial component of the account I favor is the bundle theory of persons. Thus, I take Aune to be a bundle of elements. Among these elements is the red expanse.

For the red expanse to be before Aune's mind just is for it to be a part of the bundle he is. Put more generally—to have a quale is for the quale to be a member of the bundle you are.

The picture that emerges is this: The qualia you have are the basic reality designated by the neurophysiological concepts which the neurologist applies to your brain. Put less formally—qualia are (a part of) what the brain is made of. Qualia also are among the elements that compose the person whose brain they form. A person has exactly those qualia that feature in the bundle she is.

1.3 Outline Two: Chapter Summaries

So much for a very brief sketch of the sort of view I want to defend and the sorts of considerations that lead me in that direction. I will now traverse the same territory once more by way of a very brief summary of each chapter.[5]

Chapter One, Chapter Two, and Chapter Three. These chapters are concerned with introductory and methodological matters. In the two chapters following this brief introductory chapter I clarify my terminology and defend the methodological choice of approaching consciousness from the first-person point of view.

Chapter Two. In this chapter I describe qualitative consciousness and distinguish it from other forms of consciousness. The notion of a quale is introduced in an intuitive manner. And I attempt to evoke a lively appreciation of the existence and the importance of the problems posed by the existence of qualitative consciousness.

Chapter Three. Chapter Three opens with a statement of the principle of phenomenological adequacy: a theory of consciousness must preserve the phenomenological data afforded us by our unique first-person point of view on consciousness. I defend the first-person point of view against all sorts of accusations: that it is a form of obscurantism; that it is nothing but covert dualism; that it fosters an anti-scientific attitude, etc. The most serious challenge to the first-person point of view is Daniel Dennett's method of 'heterophenomenology.' Dennett argues that his method has all the advantages and none of the problems of its more traditional rival. I try to show that heterophenomenology does not hold what it promises. Some reflections on the place of this project within the wider field of the philosophy of mind bring this chapter to an end. It will become clear how narrow and how important the present inquiry is.

Chapter Four, Chapter Five, and Chapter Six. That introspection holds the key to consciousness is an idea that has and still does inform much thought about consciousness. In these three chapters I present and criticize current introspectionist accounts of how we have our qualia.

Chapter Four. Theories that attempt to explain consciousness in terms of higher-order representations have played a prominent role in recent discussions of consciousness. The proponents of these views have tended to emphasize the differences between their accounts, and the differences are considerable. Some hold that qualia are external, physical properties. Others hold that qualia are internal, neurophysiological properties. Some hold that higher-order representation is perceptual. Others hold that it is purely cognitive. In this chapter I present a unified view of the various higher-order representation accounts. I see them all as minor variations on the traditional introspectionist accounts of consciousness. These accounts are the topic of the next two chapters. Along the way I refine the notion of a quale by drawing a number of distinctions that were not apparent in Chapter Two. These distinctions are used throughout the rest of the discussion.

Chapter Five. To many philosophers it has seemed *obvious* that introspection holds the key to consciousness. In this chapter I try to expose and criticize the hidden causes of this tendency towards introspectionism. My aim is to show that introspectionism is unsupported. This falls short of proving introspectionism false. But if I succeed in changing the attitude towards introspectionism from 'It's obvious' to 'It looks quite unsupported' I would count this as a success.

Chapter Six. Oscar is an artificial intelligence created in the image of the human mind. Oscar's consciousness is engineered according to introspectionist principles. Using Oscar as an example, I try to strengthen the conclusion reached in the previous chapter. Introspectionism is unsupported and in addition there are good reasons for thinking it false.

Chapter Seven, Chapter Eight, and Chapter Nine. The traditional notion of a quale makes qualia into Mental, i.e., nonphysical properties designed to be borne by nonphysical sense-data. Relocated qualia are qualia that have been moved out of the Mind and onto some nonMental bearer. In these three chapters I discuss the promises and problems of the relocation move.

Chapter Seven. I begin by discussing various ways of deMenting qualia—ways, that is, of moving qualia out of the Mind—and the reasons for undertaking this move. Two versions of the relocation thesis are distinguished and assessed. Wholehearted relocation moves the (color) qualia all the way out onto the surface of the perceived physical object. Halfhearted relocation moves the qualia out of the Mind and into the brain.

Chapter Eight. All the representation-based accounts of consciousness I discuss employ deMented relocated qualia. In this chapter I present a typology of four distinct types of representation-based accounts of consciousness. I arrive at the conclusion that having relocated qualia is not a matter of representing them. The difficulties encountered by the several types of representation-based ac-

counts vary. But the common denominator is their failure to live up to the standard of phenomenological adequacy.

Chapter Nine. Nonveridical experiences—illusions, hallucinations, etc.—pose a well-known problem for wholehearted relocationists. For in these cases there exists no external, physical object to bear the experienced qualia. The traditional answer to this problem has it that in these cases the qualia are borne by merely intentional objects. I take this answer at face value and examine a number of construals of merely intentional objects to see whether they would make a promising home for qualia. The results are not encouraging. I then turn my attention to the much more important domain of veridical experience. I try to argue that the same problems that arose in the domain of nonveridical experience resurface in the domain of veridical experience. The halfhearted relocationist seems to fare much better vis-à-vis nonveridical experience. But as regards veridical experience halfhearted relocationism is no better off than wholehearted relocationism. Therefore, the relocation of qualia must be avoided.

Chapter Ten. A survey of the problems encountered so far suggests that a successful theory of consciousness should be guided by two maxims. First, experience must not be analyzed on the act/object model, according to which the act is that which lets one have the quale exemplified by the object. Our survey has shown that none of the proffered act candidates can do their job. Therefore, we need an account of experience that lets one have one's qualia in a manner that is not mediated by an act. In short, we need a nonrelational theory of experience. Second, qualia must not be understood as attributes of underlying substances. Throughout we have assumed that there are no Mental qualia-bearers. And in Chapter Eight and Chapter Nine we have seen that one cannot make a theory of consciousness out of the sorts of qualia that nonMental qualia-bearers can bear. This means that the sorts of qualia we need cannot be borne by anything Mental or by anything nonMental. Hence, if qualia (of the relevant sort) need bearers to be instantiated, then there are no qualia. But there are qualia. The only way we can have qualia without bearers is to make them into properties that need no bearers in order to be instantiated.

I discuss two theories of consciousness that heed these maxims: adverbialism and the view I call 'monism.' Adverbialism is elegant but fails to satisfy the condition of phenomenological adequacy. Monism is markedly inelegant, yet it appears to satisfy all the conditions on an acceptable theory of consciousness. The monist argues that a person is a bundle that has a quale just in case the quale is an element in the bundle the person is. And the monist regards qualia as irreducible features that fit into the physical world as the underlying reality of the organ the neurophysiologist knows as the brain. Taken together, the monistic

answer to the having-problem and to the nature-problem comprise the monistic theory of qualitative consciousness.

The monist's answers to the having-problem and to the nature-problem raise many questions. Two of these questions are discussed at some length. First, it may seem that the monistic account of how qualia are had makes us infallible about qualia. So, since we are fallible about these matters, monism is false. I reply by arguing that monism is a noncognitive theory of consciousness. According to monism, having a quale is not a matter of cognizing anything in any way. So monism does not entail infallibilism. Second, it may seem that the monistic account of the nature-problem entails panpsychism. And that, many would say, reduces monism to absurdity. I reply by arguing that the case for panpsychism is weaker than it seems on first sight. And the end of the world would not be upon us even if the attempt to reign in the psyche should fail. For the version of panpsychism that monism may entail lacks the features that make it repugnant to the naturalist.

1.4 Caveats

On page 5 I listed three questions that we should ask of any purported theory (T) of consciousness:

1. Do the qualia of T fit into a naturalistically conceived world?
2. Does the having-relation of T fit into a naturalistically conceived world?
3. Does the fact that a person has (in this manner) a quale (of this sort) make it clear why being so related to such a thing results in its being like something to be that person?

Many authors seem to be in the enviable position of being able to answer Yes!, Yes!, and Yes! when asking these questions of their own theories. I barely manage to answer Yes?, Yes?, Yes?. I hope that this is less a function of the intrinsic unbelievability of what I say than an expression of a general doubt about what philosophy can establish in such matters. Russell's view of this issue accurately captures my feelings about the outcome of the present investigation:

> What will have happened will be that we shall have come to see a complicated structure where we thought everything was simple, that we shall have become aware of the penumbra of uncertainty surrounding the situations which inspire no doubt, that we shall find doubt more frequently justified than we supposed, and that even the most plausible premisses will have shown themselves capable of yielding unplausible conclusions. The net result is to substitute articulate hesitation for inarticulate certainty. (Russell 1980a: 11)

I find myself deeply convinced that being conscious is a matter of having qualia. But as the complicated structures pile up in the process of articulating this feeling, conviction pales and hesitation takes over. Long years of intense hesitation have taught me that it is not a very good method to write books by. Whether the method of asserting with apparent conviction what one hardly dares to believe is any better, the reader may judge for himself or herself.

As for articulateness. If one German as mother tongue has, then tends one towards English sentences wrong way around to write. I can only hope that you will find the many stylistic infelicities amusing and I trust that you will take it for granted that anything that seems obscure, unclear, or worse in these pages is due solely to my linguistic handicap.

CHAPTER TWO

Making our Ideas Clear: Consciousness and Qualia

2.1 The Thesis

This is an essay about qualia and consciousness. The main thesis I wish to defend concerns the relation between qualia and consciousness. I argue that *to be conscious is to have qualia.* This thesis does, I fear, encapsulate everything of importance that I have to say about qualia and consciousness. But despite its seeming simplicity, this thesis bears explanation. The task I have set myself is to transform this slogan into a substantial philosophical thesis.

2.2 Clarifying the Thesis

Without further elucidation the meaning of the slogan 'to be conscious is to have qualia' is so vague as to be virtually empty—the analysandum is unclear and the analysans is cryptic. The point of an explanation, explication, or analysis of a concept is to get the concept into sharper focus. The complaint that the analysandum is unclear may therefore seem out of place. The analysis will provide the necessary clarification. To ask for a clarification of the analysandum before beginning the analysis is to ask for an analysis before the analysis, which seems absurd. But this is only an apparent paradox. Before we can sensibly embark on a project of conceptual analysis or explication we need to have a rough understanding of the analysandum or explicandum. The term 'consciousness' is sufficiently ambiguous to require such preliminary regimentation.

The proposed analysans—the having of qualia—is in worse shape than the analysandum. It is not clear what we mean by the term 'consciousness'; but few have doubted that we are conscious. In the case of qualia, even this minimal consensus is absent. The meaning of the term 'quale' (pl. 'qualia') is unclear and many have doubted that there are qualia. Therefore, this investigation must begin with a sketch of the relevant notion of consciousness followed by an apology for

qualia. By the end of these preliminary considerations, the thesis that being conscious is a matter of having qualia will, hopefully, have become clear enough to merit further discussion.

2.2.1 Qualitative Consciousness

Even if we stick to what one might call the 'philosophical' uses of the term 'consciousness,' we find a great deal of disagreement. I will not attempt to determine what are the right or wrong uses of this term. I simply want to state how I am going to use the term 'consciousness.' The use to which I put the term 'consciousness' is neither new nor unusual.[6]

Thomas Nagel has provided the most succinct characterization of this notion of consciousness when he wrote:

> the fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism...fundamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism. (Nagel 1974: 166)[7]

This formulation points directly at the salient but hard to describe phenomenon that is qualitative consciousness. To many of Nagel's readers the 'what it's like' locution has seemed intuitively obvious and brilliantly illuminating. But it would be disingenuous to pretend that Nagel's 'what it's like' formula has found universal acclaim—far from it.[8] Some philosophers were puzzled by the Nagelian phrase. Others have held it to be incomprehensible, misleading, trivial, or whatnot. I have no hope of persuading everyone of the merits of the 'what it's like' locution. But I would like to briefly address four sources of possible misunderstanding.

First, a number of philosophers insists on reading the 'what it's like' locution in a comparative sense, notwithstanding Nagel's explicit admonition to the contrary. On this reading the expression 'what it's like' means the same as the expression "what (in our experience) it resembles." (See Nagel (1974: 170)) As this is not the sense of the formula that Nagel and those who follow him in finding his formula illuminating intend I will say no more about it.

Second, a good part of the criticism aimed at the 'what it's like' locution is really directed at a popular variant of it: *knowing* what it's like. The knowledge claim contained in this version of the Nagelian catchphrase has attracted much critical scrutiny. Since I am not concerned with someone's *knowing* what it is like, but merely with the fact of its being like something for someone, I will not address these criticisms.[9] But I will take time to briefly consider why *knowing* what it's like has attracted so much attention.

Reasons for this shift of focus towards the epistemic reading of Nagel's original formula are not hard to find. Nagel's discussion of the nature of a bat's experience is one of the sources facilitating the slide from the expression 'what it's like' to the expression 'knowing what it's like.' The point that many readers glean from Nagel's discussion is epistemic: we cannot know what it's like to be a bat. Thus *knowing* what it's like is now at center stage. That Nagel's discussion of the bat should give rise to this slippage is doubly ironic. For, first, though very vivid, the discussion of the bat is purely illustrative of a certain component of Nagel's argument—an argument that could stand without making any mention of bats or other strange creatures. Second, as in the previous point, Nagel has explicitly repudiated the epistemic reading of his bat example: "My point, however, is not that we cannot *know* what it is like to be a bat. I am not raising that epistemological problem." (Nagel 1974: 172) Notwithstanding all that, the bat is still to blame for the widespread preoccupation with *knowing* what it's like. An additional source predisposing one to fixate on the epistemic version of Nagel's formula is a tendency to run Nagel's argument together with Frank Jackson's knowledge argument. That done, *knowing* what it's like must seem like the natural focus of Nagel's worries.

Third, it has been argued that "it is unclear that the formula identifies conscious states as such, since one might want to say there was something it was like to be a stone, e.g. grey, detached, and not all that heavy." (Hannay 1990: 186) How should one reply? One cannot deny that the stone is like something, for it is grey, detached, and light, etc. But to acknowledge that the stone is like something is not the same as to grant that it is like something for the stone to be a stone. From the 'stone's point of view' things don't appear in any way at all. The stone's being grey does not translate into its being like something to be grey for the stone. None of the stone's properties are reflected in its being like something for the stone to have these properties. That is to say, the stone is not conscious. Thus, I would try to meet this objection by granting the point it raises, but by insisting that it does not succeed in undermining the proposed attempt to delimit a certain interesting notion of consciousness. Everything is like something but only conscious things are such that it is like something for them to be them.

A fourth point. It is like something to experience a stone. And the experienced stone is like something. But it is a mistake to infer that the 'what it's like' locution fails to pick out the characteristic feature of consciousness on the grounds that the stone too has the feature the locution picks out. This is the point I just made—a point that appears to belabor the obvious. T.L.S. Sprigge has generalized this observation in an ingenious and nontrivial way. The stone does not have the relevant what-it's-like dimension. That is obvious. And it may seem no

less obvious what the true bearer of the what-it's-like dimension is: it is the experience of the stone. But this too is a mistake according to Sprigge:

> The formula concerns publicly observable individuals, not experiences. It is the claim that for an individual to be conscious, there must be something that it is like to be him, her, or it...It is not the claim (with which it is sometimes confused) that there is something that a *conscious experience* is like...That is quite an unhelpful suggestion, since there is nothing such that it is not like something (either in the sense of being similar to something else or, more relevantly, in having a character). The point of the formula is to distinguish things concerning which there are *also* truths as to what it is like being them. (Sprigge 1994: 79-80)

It is instructive to consider Sprigge's observation on the background of the state consciousness/creature consciousness distinction. David Rosenthal, the originator of this widely accepted distinction, provides the following schematic characterization of creature consciousness: "To be conscious, a person or other creature must be awake and sentient." (Rosenthal fc.: Section I) He sees no deep problems here and thus focuses his attention on state consciousness: What does a mental state's being conscious consist in? A satisfactory answer to this question would provide an account of qualitative consciousness. Later on I will discuss Rosenthal's answer to this question at great length (see Chapter Four and Section 8.2). Now I merely want to warn against a possible misunderstanding suggested by the term 'state consciousness.' The feature targeted by the 'what it's like' locution is not something that a state can have 'all by itself.' The crucial feature of a conscious state is that it is like something *for the subject who has the state*. If you direct your blinkered gaze narrowly upon the state, in abstraction from the subject who has it, you will miss qualitative consciousness. This is the point that Sprigge is driving at. And the talk of 'state consciousness' has a tendency to obscure this essential 'subject involvement' of qualitative consciousness. Not the state itself but being the creature that has the state (in the relevant way) is like something.[10]

With these clarifications of the 'what it's like' locution in place, I hope to have provided a helpful pointer towards the specific notion of consciousness that is the subject of this investigation. It is this notion of qualitative consciousness (or consciousness, for short) that I propose to analyze as the having of qualia.

Nagel's characterization of qualitative consciousness is now over 20 years old. And a brief look at contemporary attempts to characterize qualitative consciousness proves that the situation is essentially unchanged after two decades.[11] At this time Ned Block is the leading advocate of the case for qualitative consciousness (or phenomenal consciousness, as he calls it). Here is how he tries to get the idea across:

> Consciousness in the sense discussed is phenomenal consciousness. 'What is *that*?', you ask. There is no non-circular definition to be offered; the best that can be done is the offering of synonyms, examples and one or another type of pointing to the phenomenon...For example I used as synonyms 'subjective experience' and 'what it is like to be us'. In explaining phenomenal consciousness, one can also appeal to conscious properties or qualities, e.g. the ways things seem to us or immediate phenomenological qualities. Or one can appeal to examples: the ways things look or sound, the way pain feels and more generally the experiential properties of sensations, feelings and perceptual experiences. (Block 1994: 210-211)

In another illuminating passage he attempts to highlight qualitative consciousness by drawing our attention to the ways in which this form of consciousness is involved in various puzzles in the philosophy of mind:

> This is the concept of consciousness that leads us to speak of an "explanatory gap": At this stage in the relevant sciences we have no idea how the neural substrate of my pain can explain why my pain feels like *this* rather than some other way or no way at all. This is the concept of consciousness that gives rise to the famous "inverted spectrum" hypothesis—things we both call "green" look to you the way things we both call "red" look to me—and the "absent qualia" hypothesis, the idea that there could be a machine that was computationally like us, but was nonetheless a phenomenally unconscious zombie. (Block 1992: 205)

The attempt to clarify the notion of qualitative consciousness by associating it with these philosophical classics is helpful but risky. For there are always those who merely see a pseudo problem where the rest of us perceive an uncomfortable paradox. And to them it may appear that qualitative consciousness is just as pseudo as the puzzles it allegedly creates. Block has seen this risk. His comment on the puzzles-based way of explaining qualitative consciousness puts this matter into clearer perspective:

> Note that these conundra are *routes* to phenomenal consciousness—they do not constitute it. One can accept phenomenal consciousness without accepting any of them because our fundamental access to phenomenal consciousness derives from our acquaintance with it. (Block 1992: 205)[12]

Another route to qualitative consciousness is provided by contrastive analysis. We can attempt to further illuminate the nature of qualitative consciousness by contrasting it with a number of strikingly different notions of consciousness. It has been proposed that being conscious is a matter of having and exercising one of the following mechanisms or capacities:

- possessing a self-scanning or self-monitoring device (an introspector) built into one's central system.

- possessing some central control mechanism, some high level executive unit, some supreme organizer.
- being able to integrate and process information from various information sources and use this information to guide one's behavior.
- being able to be in perceptual states that represent the observable objects and states of affairs with which one's environment affects one.
- being able to be in reflective states that represent actual and possible states of affairs one contemplates.
- being able to be in introspective states that represent one's own mental states.

Numerous analyses of these (and other) mechanisms and capacities have been proposed. But invariably they fail to shed light on the question of what it is like to be the subject possessing and exercising these mechanisms and capacities. This is no accident. What these mechanisms and capacities are and what it is to have and exercise them can be fully understood from a purely third-person perspective.

The very feature that makes these other notions of consciousness so attractive to some thinkers makes them appear barren to those interested in qualitative consciousness. For what makes these other notions so much more tractable (and therefore attractive for people who like to get things done) is the fact that they can be fruitfully discussed and analyzed without having to broach the daunting problem of qualitative consciousness. But their leaving out the dimension of qualitative consciousness is, of course, precisely what makes them uninteresting to those who seek an understanding of this problematic notion of consciousness. The notion of consciousness I am interested in differs from these. It is characterized by the fact that it is conceived from the subject's point of view. But none of the listed mechanisms or capacities are so characterized.

In denying that there is a conceptual link between the notions of consciousness listed above and the notion of qualitative consciousness, I do not mean to deny all connections between the possession and exercise of these mechanisms and capacities, on the one hand, and qualitative consciousness on the other. We do possess such mechanisms and capacities. And under normal conditions it is like something for us to exercise (at least some of) these mechanisms and capacities. Take perception, for example. Normally it is like something for us to perceive our environment. But in cases of subliminal perception and blindsight one perceives one's environment without its being like anything to do so. These are cases of perception without consciousness. Thus, being conscious is not simply a matter of one's having and using one's capacity of perceiving one's environment. For the latter can (and sometimes does) occur without the former. Similar observations apply to the other proposed mechanisms and capacities

listed above. They all can (and sometimes do) get exercised without its being like anything to do so.

But perhaps this is too facile. Everybody grants that *under normal conditions* the possession and exercise of these mechanisms and capacities is accompanied by qualitative consciousness. This suggests that we say the following: Under normal conditions the possession and exercise of (at least some of) these mechanisms and capacities is causally sufficient for one to be the kind of thing it is like something to be. This seems like a reasonable claim. It will be difficult to specify the normal conditions; but that something is hard to do should not count against it. Nevertheless, I want to resist taking this path. What I want to understand is not what kind of causal machinery it takes for one to achieve consciousness.[13] I want to understand what it is for one to be conscious, what its being like something for one consists in. To this last question I want to answer that being conscious in this sense is a matter of one's having qualia. It seems, therefore, that an investigation of consciousness should not let itself be side-tracked by an investigation of these mechanisms and capacities.

Here is another way of putting the point. The practice of studying mental phenomena based on the assumption that they can be understood in abstraction from consciousness has gained sufficient prominence to earn its own label. Owen Flanagan has baptized it *conscious inessentialism* and explains it as follows: "[Conscious inessentialism] is the view that for any mental activity *M* performed in any cognitive domain *D*, even if *we* do *M* with conscious accompaniments, *M* can in principle be done without these conscious accompaniments." (Flanagan 1991: 309) I believe we can understand the view that all of consciousness is captured by the possession and exercise of the mechanisms and capacities listed above as a somewhat paradoxical manifestation of a pervasive conscious inessentialism. While I do not mean to be critical of conscious inessentialism in general, I do believe that this method has its limits. In particular, we cannot hope to achieve an understanding of consciousness by employing a method that advises us to regard consciousness as not essential to what we study. For when we study consciousness, consciousness is, obviously, essential to what we study.

2.2.2 An Apology for Qualia

2.2.2.1 The Datum

At every moment of our waking and dreaming lives we are immersed in a sea of colors, sounds, smells, tastes, tickles, pains, and the like. Throughout our lives we are subject to a ceaseless barrage of such qualities. At every moment an abundance of such qualities confronts us. Nothing seems more obvious to me.

Nothing seems more important to me. And nothing seems more baffling to me. Let me explain.

2.2.2.2 Why the Datum Is Obvious

That we are confronted (at every moment of our waking and dreaming lives) with colors, smells, tastes, etc. is not something I hold as a result of an argument. Every justificatory argument that I might invent for this conclusion would suffer from the defect that its premises would be much more questionable than the conclusion it aims to establish. I am, therefore, in no position rationally to convince a dissenter of the truth of my belief. Nor do I know how to 'show' the dissenter the truth of my belief by exposing him to novel situations, by training him, or by teaching him in any other way. The way in which we are inundated by these qualities is so pervasive and so blatant that any attempt to draw a person's attention to it must appear comical or absurd. The phenomenon I am talking about is not something subtle or refined or recondite. To appreciate it you need not ingest any drugs, fast, exercise any form of monastic self-discipline, or become a member of any particular philosophical sect. To appreciate the phenomenon in question you need not think at all. It takes no focused attention, no process of abstraction or distillation, no cultivation of any form of cognitive refinement whatsoever to discover that we are swamped by a multiplicity of colors, sounds, tastes, and bodily feelings at every moment of our waking and dreaming lives. This is a datum that stands fast for us.

2.2.2.3 The Neutrality Thesis

There are those who appear to dissent. This may seem astonishing. But this impression lasts only as long as we fail to notice that the disagreement is about the philosophical interpretation of the datum in question. Thus, the dispute will not be settled by more and more appeals to the obvious. We may, however, be able to appease most dissenters by providing a philosophically innocuous interpretation of the datum that stands fast for us.

The trouble starts with the terminology. How is one to describe the datum without offending certain metaphysical sensibilities right there? The reds, the blues, the pancake smells, and the thunder claps that we meet with in experience have received more than their fair share of impressive sounding names. Generally, these kinds of items have been understood as qualities, or properties, or features, or characters whose differentia consisted in their being sensible, or phenomenal, or phenomenological, or qualitative, or subjective, or felt. And the particulars graced by these items were usually taken to be phenomenal objects like sense-data, or sensations, or experiences, or ideas, or Mental objects of some other sort. One can readily see how to generate a large number of names

(nearly all of which can be found in the literature) from these three lists of terms: 'the subjective character of experience,' 'the sensible properties of sense-data,' 'the phenomenal qualities of sensations,' and so on. In addition, we find a number of evocative expressions whose purpose it is to describe the effect of our being immersed in all these phenomenal qualities, and in particular our being immersed in the multitude of colors. Here we find 'the inner light,' 'the inner illumination,' 'the inner glow,' 'the technicolor phenomenology,' 'mental paint,' or simply 'phenomenology.' And finally, there are two technical terms to be considered. Herbert Feigl's 'raw feels' and William James's 'qualia.'

All of these many expressions share one unfortunate feature. They all suggest, more or less strongly, that the colors, sounds, tastes, etc., in which we find ourselves immersed are Mental in the sense that they qualify some sort of inner Mental object. This has the following consequence. If we simply describe our original datum in traditional philosophical terms, we end up making a metaphysically loaded and materialistically unacceptable claim entailing the existence of phenomenal individuals. A skeptical attitude towards such peculiarly Mental objects will therefore manifest itself as a skepticism about the existence of our datum. This, then, is one reason why somebody might want to deny the obvious datum.

This dispute is easily avoided by simply stipulating that we are going to use the problematic terms without any metaphysical strings attached. I propose to use the terms 'quale' and 'qualia' (and all the other terms listed above) in such a way that the question of the nature of the qualified object remains open. The quale might qualify the external physical object I see, the sense-datum I have, my present experience, the brain state that is, or is involved in, my seeing something, or whatnot. This, then, is the sense in which qualia are, for the moment, *metaphysically neutral*—neutral, that is, as between the Mental/physical distinction. In being confronted with a quale, I am confronted with a property of something; but at this point we will leave the question of the nature and the location of the qualia-bearing object open.

It is a mistake to start off by bifurcating properties into the objective properties of external objects and the subjective qualities of the mind. We may eventually be forced to adopt such a two-tiered account. But we should not start off by assuming that qualia belong to the subjective pole of the spectrum. The red quale that I am confronted with upon viewing a ripe tomato may simply be one of the properties of the tomato itself, rather than an intrinsic quality of my sensing the tomato. This should neutralize what is, perhaps, the main reason for rejecting the initial datum.[14]

2.2.2.4 Qualia and Sense-Data

Some of the ill will against qualia may stem from the feeling that the talk about qualia is nothing but thinly veiled talk about sense-data. And sense-data, so the present philosophical orthodoxy would have it, are unacceptable. I will not discuss the reasons for the widespread rejection of sense-data. They seem to me rather less convincing than the virtually unanimous rejection of sense-data would lead one to expect.[15] Be this as it may, the suspicion that qualia are just resurrected sense-data is quite mistaken. The more careful discussions of sense-data make clear that sense-data are introduced as bearers of qualia. John Wisdom states the point clearly: "Sense data are not identical with sense-qualities but are the things which possess *or* seem to possess sense-qualities." (Wisdom 1963: 142) This very distinction has played a major role in the thinking of the most meticulous of all sense-data philosophers. Commenting critically on his earlier view that the color, size, and shape of a visually perceived patch are sense-data, G.E. Moore has this to say: "I should now make, and have for many years made, a sharp distinction between what I have called the "patch," on the one hand, and the color, size and shape, *of* which it is, on the other; and should call, and have called, *only* the patch, *not* its color, size or shape, a "sense-datum."" (Moore 1966: 44, fn. 2) Moore distinguishes between the sensible qualities, like a particular color, size, and shape, and the sense-datum that has those qualities, like a patch, a line, a spot, or a speck. This is not merely some trifling afterthought that the sense-data philosopher could do without. As Moore's discussion of adverbialism makes clear, the insistence on the distinction between sense-data and their sensible qualities is the crucial point of the sense-datum theorist's defense of his theory against its most serious competitor. (See Schilpp (1968: 659-660))

This brief discussion of sense-data should lay to rest all worries about qualia that stem from a confusion between qualia and sense-data. The acknowledgment of metaphysically neutral qualia does not commit one to murky sense-data. This consideration should remove a second powerful motivation for denying our initial datum.

2.2.2.5 The Qualia Quiners

The preceding considerations should have removed some doubts about the genuineness of the initial datum to which I appealed above. By disentangling qualia from the suspect sense-data and by proposing to understand qualia as metaphysically neutral, I hope to have removed the major causes for one's wanting to resist the claim that we are immersed in a sea of qualia.

But opposition to the obvious datum may stem from sources that are less easily neutralized. Take, for example, Daniel Dennett's indictment of qualia:

"'Qualia' is a philosopher's term which fosters nothing but confusion, and refers in the end to no properties or features at all." (Dennett 1988: 49) The cheap reply to Dennett's eliminativism is to simply note that his notion of qualia differs from the one I have been proposing here. The properties Dennett wants to eliminate are philosophically distilled properties of conscious experience whose special nature consists in their fourfold essence: ineffability, intrinsicalness, privateness, and direct apprehensibility. Dennett's notion of a quale is the notion of a special mental property that is arrived at through some sophisticated process of abstraction. I, on the other hand, mean to talk about thoroughly unremarkable metaphysically neutral properties that force themselves upon one without the need for any distillation process. Since the properties that Dennett wants to eliminate are quite different from the one's the existence of which I take to be obvious, I might simply shrug off Dennett's eliminativism as irrelevant.[16]

This dismissal of the qualia Quiners may be a little bit too facile. First, the inoffensive notion of qualia in terms of which I want to state the initial datum will, in all likelihood, not be the notion of qualia with which this investigation ends. And the more refined notion of a quale that emerges in the course of the investigation may turn out to be uncomfortably close to the notion of qualia that Dennett is attacking. Second, the arguments upon which Dennett bases his qualia eliminativism may be strong enough to cast doubt even on the relatively harmless notion of qualia with which I want to start out initially. We will, therefore, be well advised to examine Dennett's reasons for his skepticism about qualia. If we can find a flaw in his argument the above worries need no longer trouble us.

Dennett's argument proceeds by confronting us with numerous plausible thought experiments ('intuition pumps') all of which are variations on a common theme. Dennett makes a surprisingly plausible case for the claim that re-identifying qualia is not the simple task we may have taken it to be. Consider three of the puzzling cases presented in Dennett (1988). The disenchanted coffee drinker used to like coffee but now she no longer enjoys it. Is this so because the way the coffee tastes to her has changed? Or have her tastes changed resulting in her no longer liking the very same taste quale? The experienced beer drinker finds himself in a similar predicament. Now he loves the taste of beer; but many years ago, when he first tasted beer, it tasted horrible. Is this a matter of developing a liking for the same taste? Or is the experienced beer drinker better understood as someone for whom the taste of beer has gradually changed? Finally, there is the case of the cauliflower conversion. A confirmed cauliflower hater, you find yourself loving the taste of cauliflower after taking the miraculous pill. How has the pill done its work? Has it changed your qualia, leaving your likes and dislikes untouched? Or has it changed your attitudes towards the same old cauliflower qualia? An appeal to experience cannot settle the question: the cauli-

flower experience as a whole is now so thoroughly transformed that it resists a neat analysis into one part that stayed identically the same and another part that changed. In these and similar cases we do not seem to have a principled way to tell whether the qualia did or did not stay the same. We have to make what looks like an arbitrary choice. These quandaries are not an artifact of Dennett's particular notion of qualia; the same problems arise even on the most unexacting notion of qualia.

What does reflection upon these examples teach us? The lesson that Dennett wants to draw from the consideration of such cases is momentous. The elusiveness of the purported qualia, the difficulty of telling when we are dealing with the same ones again, persuades Dennett to deny their existence. Positing an elusive quale at the core of the various reactive attitudes that characterize the experience adds nothing to the experience. We can go with the reactive attitudes alone. The quale itself is a superfluous fiction.[17]

Can one accept the cogency of Dennett's intuition pumps and, at the same time, resist Dennett's eliminativist conclusion? I think this is possible. For I believe that Dennett's qualia eliminativism is not a lesson innocently gleaned from these examples; it is a philosophically loaded and controversial interpretation of his ingenious puzzle cases concerning the identity of qualia. Dennett's reasoning is based on two principles: the principle of first-person operationalism, and the principle that found its expression in the slogan 'No entity without identity.'

First-person operationalism "denies the possibility in principle of consciousness of a stimulus in the absence of the subject's belief in that consciousness." (Dennett 1991: 132) In other words, differences in one's qualitative consciousness cannot outrun differences in one's beliefs about it. Applied to the cases of the coffee drinker, the beer drinker, and the cauliflower hater the principle rules out what otherwise would have seemed like the obvious reply to these puzzles. Concerning the experiences of the veteran coffee drinker, for example, one might be inclined to say this. It is possible that she does not believe that her qualia changed (and her tastes stayed the same) and that she also does not believe that her qualia stayed the same (and her tastes changed). But her failure to have determinate beliefs about the sameness or difference of her qualia does not show that there is anything indeterminate about her qualia. Either her later qualia are the same as her earlier qualia or they are not; her beliefs about this matter are irrelevant. On this interpretation, the case of the coffee drinker shows only that we know our qualia imperfectly, not that our qualia are indeterminate. To draw a stronger conclusion would be to confuse ignorance *of* qualia with indeterminacy *in* qualia. But first-person operationalism rules out this answer. The coffee drinker's qualitative consciousness cannot be more determinate than her beliefs

about it. Therefore, it is indeterminate whether her present and her previous qualia are the same.

In this way first-person operationalism sets the stage for the 'No entity without identity' principle. The latter principle licenses the inference from the indeterminateness of questions about the identity of qualia to the conclusion that there are no qualia. In this way the joint operation of first-person operationalism and the 'No entity without identity' principle lead to qualia eliminativism.

Neither one of these principles is beyond question. In Chapter Five I spend considerable time attacking the idea underlying the principle of first-person operationalism. There I try to establish that the belief thesis—the thesis that every qualitative state must be accompanied by a corresponding qualitative belief—is false (see especially Section 5.9). Thus, I reject Dennett's principle of first-person operationalism and maintain that differences in qualitative consciousness are not conditional upon the existence of corresponding beliefs. The way things seem is prior to, and independent of, our beliefs about how things seem.[18]

Rejection of the principle of first-person operationalism is sufficient to block Dennett's route to qualia eliminativism. It is, nevertheless, worthwhile to point out that the second principle Dennett relies upon—no entity without identity—is not sacrosanct either. Many kinds of entities confront us with seemingly irresoluble puzzles about identity. In all these cases we have to make a choice: either give up these entities or give up the metaphysical maxim. I suggest that in the case of qualia we may want to give up the maxim.[19] These considerations persuade me that the challenge of qualia eliminativism can be set aside. It still remains to be established that to be conscious is to have qualia. But at least we need not fear that this analysis will fail on the grounds that there are no qualia.

2.2.2.6 Why the Datum Is Important

Recall the quotation at the beginning of the introduction: "Why, if it weren't for this 'internal illumination'…the world would be nothing but a pile of dirt!" (Feigl 1967: 138) Einstein's dictum perfectly captures my feeling about this matter. That I am confronted (at every moment of my waking and dreaming life) with colors, smells, tastes, etc. is something that is of the utmost importance to me. It is in virtue of this fact that it is like something to be me. Were it not for this confrontation with qualia my life would be like an unbroken stretch of dreamless sleep. Without my qualia I am, as far as *I* am concerned, dead.[20]

This assessment of the role of qualia for being myself to myself may seem exaggerated. Would not the 'dequaliafied' person at least be capable of enjoying 'the life of the mind'? One might think that the state of a 'dequaliafied' person is similar to the state that St. Thomas ascribes to disembodied souls: they do not see, hear, smell, etc.; nor do they feel pain, anger or fear. But they can think and

will. (See Geach (1969)) But P.T. Geach argues persuasively that being such a nonsensuous intellect is a rather sorry affair. His doubts that such 'dehumanized thinking and willing' is possible rest on the following consideration:

> In our human life thinking and choosing are intricately bound up with a play of sensations and mental images and emotions; if after a lifetime of thinking and choosing in this human way there is left only a disembodied mind whose thought is wholly non-sensuous and whose rational choices are unaccompanied by any human feelings—can we still say there remains the same person? Surely not: such a soul is not the person who died but a mere remnant of him. (Geach 1969: 22)

The view that qualia are not that central a feature of our being is grounded in a mistaken view about the scope of qualia. Even astute qualophiles (Dennett's term) tend to limit the range of qualia to the domain of secondary qualities. I suspect that this narrow view about the extent of qualia can be traced back to a particular way of introducing qualia. Limiting the range of qualia to the domain of secondary qualities will seem natural if one conceives of qualia as the solution to the problem of the physical irreducibility of secondary qualities.

But if one thinks of qualia along the lines that I have proposed then limiting qualia to secondary qualities will seem entirely baseless. The sea of qualities into which we find ourselves thrown contains shape, size, solidity, and motion no less than color, sound, and taste. And this list still does not exhaust the full range of qualities with which we are confronted. There are the pains, tickles, and all the more indistinct bodily sensations. Furthermore, there is the domain of what we may call nonsensuous phenomenology. In this area we find the qualities that go with emotions like anger and fear and with those feelings that indicate occurrent propositional attitudes like belief, hope, and doubt. Bertrand Russell, for example, distinguishes three kinds of distinct belief feelings corresponding to memory, expectation, and bare assent.[21] (See Russell (1978: 250)) In the current philosophical debate appeals to nonsensuous phenomenology show up in the most diverse areas: the epistemology of memory, perception, and a priori belief (see Plantinga (1993)); the philosophy of mind and the philosophy of psychology (see Flanagan (1992); Goldman (1993b)); the philosophy of religion (see Alston (1991)). It appears, then, that the sea of qualia in which I would have you immersed is much, much deeper than suspected. Even granting that some of the claims made on behalf of nonsensory qualia are a little too ambitious, we still have ample evidence that the realm of qualia extends far beyond the initial core of examples limited to colors, sounds, tastes, odors, and feels.

Once one appreciates the full scope of qualia, the claim that it would be like nothing to live a life without them will, I hope, be more plausible. 'The life of the mind' is parasitic upon the 'internal illumination.' If the inner light goes out

you are, in Einstein's words, nothing but a pile of dirt. Wilfrid Sellars is right: qualia are what make life worth living. (Quoted in Dennett (1991: 383)) This is why our original datum is so important.

2.2.2.7 Why the Datum Is Baffling

How can the world serve up the colors, sounds, tastes, smells, pains, and so on in which we find ourselves immersed? How do the qualia fit into a world composed of atoms and the void? This is a genuine question, not an argument. But it is a question that philosophical reflection forces upon us so irresistibly that no metaphysics that leaves it unanswered can hope to satisfy. Once our prephilosophical naive realism has been dealt its first blows, neither the physical objects around us, nor the brain in us, seem capable of sustaining qualia. Wittgenstein expressed this bafflement—the genuineness of which he denied for reasons that are hard to discern—in the most urgent terms when he spoke of the giddiness that befalls us when we face the "unbridgeable gulf between consciousness and brain-processes." The giddiness is felt most vividly when I "turn my attention in a particular way on to my own consciousness, and, astonished, say to myself: THIS is supposed to be produced by a process in the brain!—as it were clutching my forehead." (Wittgenstein 1974: Section 412) It takes little reflection to experience the bafflement over the state of affairs registered in our initial datum. I will, nevertheless, quote a number of the more recent and most eloquent expressions of this bafflement. The point needs to be rubbed in. For I believe that keeping the amazement at our original datum alive is, perhaps, the single most important exercise for one who wants to resist being lulled into accepting a facile pseudo answer to the question of our place in nature. Here, then, are some of the expressions of the bafflement:

> We do not know how consciousness might have arisen by natural processes from antecedently existing material things. Somehow or other sentience sprang from pulpy matter, giving matter an inner aspect, but we have no idea how this leap was propelled. It seems abrupt, unmediated, serendipitous. One is tempted, however reluctantly, to turn to divine assistance: for only a kind of miracle could produce *this* from *that*. It would take a supernatural magician to extract consciousness from matter, even living matter. Consciousness appears to introduce a sharp break in the natural order—a point at which scientific naturalism runs out of steam. It has the look of an irresoluble enigma. (McGinn 1991: 45)

> How is it possible for conscious states to depend upon brain states? How can technicolor phenomenology arise from soggy grey matter? What makes the bodily organ we call the brain so radically different from other bodily organs, say the kidneys—the body parts without a trace of consciousness? How could the aggregation of millions of individually insentient neurons generate

subjective awareness? We know that brains are the *de facto* causal basis of consciousness, but we have, it seems, no understanding whatever of how this can be so. It strikes us as miraculous, eerie, even faintly comic. Somehow, we feel, the water of the physical brain is turned into the wine of consciousness, but we draw a total blank on the nature of this conversion. Neural transmissions just seem like the wrong kind of materials with which to bring consciousness into the world, but it appears that in some way they perform this mysterious feat. (McGinn 1989: 349)

There is no more amazing and puzzling fact than that of consciousness...For, as Thomas Nagel has noted, it is consciousness that makes the mind-body problem intractable; it is the unfathomable gap between physical process and subjective awareness which mocks our search for the filaments that bind the corporeal and the mental together. (Maudlin 1989: 407)

Regarded from the standpoint of physical science, the most puzzling thing about consciousness (or awareness or sentience: I use these terms as synonymous) is the fact that it exists at all. There is on the face of it absolutely nothing in the laws of physics and chemistry, as currently understood, that is capable of accounting for the extraordinary capacity of a lump of matter that we call the brain—once likened by Alan Turing to 'a lump of cold porridge'—to sustain an 'inner life.' (Lockwood 1989: 1)

Michael Lockwood has put his finger on the source of this puzzlement when he says: "if materialism is true we should...expect some kind of 'fit' between the phenomenal character of our experiences and the physical character of the goings-on in our brain which are constitutive of our having these experiences." (Lockwood 1989: 16) The baffled ones are struck by what they perceive to be a complete absence of fit between brain processes on the one hand and qualia on the other. And to some it has even seemed that the prospects of achieving a fit get worse the more we learn about the brain: "Advanced neurophysiological theory seems only to deepen the miracle." (McGinn 1989: 359)

How the fit should manifest itself is not an easy question to answer. Lockwood himself raises the provocative question: "What would consciousness have felt like if it had *felt* like billions of tiny atoms wiggling in place?" (Lockwood 1989: 16) One way in which the worry about the lack of fit has found a less metaphorical expression is Wilfrid Sellars's grain argument. We will come back to these issues in Chapter Nine.

2.2.2.8 Deflating the Bafflement

The seriousness of the bafflement might be doubted on the grounds that it might well be a trivial consequence of the manner in which we defined our terms. If one thinks of qualia as special Mental properties that cannot be had or produced by a purely physical process then their 'otherness' comes as no surprise. Hence

the bafflement would not open a door onto a deep metaphysical problem; it would simply be a case of one's getting caught in a trap one set oneself. But we avoided this trivialization of the bafflement by introducing the notion of qualia in a metaphysically neutral way. Thus, this simple way of dispelling the bafflement is not open to us.

William Lycan has provided another deflationary analysis of the bafflement. I and the rest of the baffled crowd who bemoan the supposed lack of fit between the pahysical and the phenomenal are victims of the stereoscopic fallacy:

> We look with one eye at the brain of a subject who is having a visual sensation [e.g., an intense cyan sensation]; we see nothing but a cheesy grey mass, and we would see nothing else no matter how assiduously we scraped away at the mass with our little dowel sticks (until there was nothing left). With the other eye, so to speak, we imagine *having* the intense cyan sensation. And now stereopsis fails utterly. The two eyes *will not* focus together and make a single coherent picture. The first eye's view of the subject's brain is nothing the least bit like the second eye's having of the cyan sensation; the two views are totally incongruous. No third-person, first-eye information about the brain or about anything else could tell us what it is like, in the second eye, to have the sensation. And this seems damning to materialism. (Lycan 1990: 111)

What, exactly goes through the mind of Lycan's cross-eyed anti-materialist? Before presenting Lycan's analysis of what is wrong with the ruminations of our squinting friend, I want to offer my own account of the thoughts giving rise to the bafflement. Looking at the brain one wonders: How could all the neural (or molecular or atomic or subatomic) activity that is going on in this grey cheesy brain give rise to, and exemplify, a smooth phenomenal greenish blue of the kind that I am currently imagining. The bafflement one experiences is similar to that which one experiences upon focusing one's eyes on a blackish, soft, greasy pile of graphite and on a shining, hard, cool diamond, respectively, while wondering: How could this be the same as that? Here, too, stereopsis fails in a rather spectacular manner. In this case, however, the bafflement can eventually be dispelled by learning more about chemistry in general and allotropy in particular. But the initial bafflement is justified and beneficial. It is this kind of bafflement that drives one to learn more about nature. Eventually it is dispelled by a sizable amount of scientific knowledge that enables one to see how to close the gap between these two allotropes of carbon. We are not in a similarly fortunate position with respect to the brain/cyan bafflement. The scientific story, if it exists, has not yet been told. Just like the graphite/diamond meditation does not *show* that they are not allotropes of carbon, the brain/cyan meditation cannot *show* that there is no materialist account of qualia to be had. But, in the absence of any account of

how such a baffling fact can be true, the meditation does make materialism seem like a farfetched doctrine.[22]

Lycan's understanding of the nature of this meditation is rather different. Having given the above description of the situation that gives rise to the bafflement, he continues:

> I hope the fallacy is plain. No materialist theory of the mind has ever entailed that *watching* the grey cheesy brain of someone who is *having* an intense cyan (or whatever) sensation is qualitatively or in any other way like having that sensation oneself. Watching the brain produces grey-cheesy visual sensations in the watcher, subserved by whatever neurophysiology underlies grey-cheesy visual sensations, and grey-cheesy sensations are of course *neurophysiologically* and functionally quite unlike the intense deep cyan sensations being had by the subject whose neurophysiology is being watched. No materialist has ever thought of claiming that grey cheesy sensations are phenomenologically *or* neurophysiologically like intense deep cyan ones, and no materialist view has ever entailed or even suggested such a claim. (Lycan 1990: 111)

I agree that materialism does not entail that looking at the grey brain of one who has a cyan quale affords a cyan quale to the onlooker. Materialism does not entail that, nor does any other doctrine that is not patently false. If the brain you watch is grey all over it will cause you to have a grey quale. And since a grey quale is quite unlike a cyan quale, nobody in their right mind would expect that watching a grey brain would be anything like experiencing a cyan quale. A grey quale is as different as could be from a cyan quale. This is obvious and has baffled nobody. If someone professed to be baffled by this fact, he would be making a silly mistake not deserving the honorific title 'fallacy.' What baffles is that any kind of neurophysiological process could produce any kind of quale at all, be it grey, cyan, or whatever. The fact that the brain in the observer's head is serving up a grey cheesy quale is no less astonishing than the fact that the grey, cheesy, brain that is being observed manages to serve up a cyan quale to the subject whose brain it is. This sort of bafflement is likely to become particularly acute when we actually look at a brain; not because the brain is grey and cheesy to look at, but because it seems unintelligible how the kinds of processes going on in a brain could give rise to and could instantiate any color qualia whatsoever.[23]

One more comment on Lycan before I move on. I have argued that the baffled ones will not and should not be persuaded by Lycan's deflationary analysis of their bewilderment. This criticism notwithstanding, I do believe that Lycan's discussion of the stereoscopic fallacy does point towards a disturbing and important problem. If we assume that qualia are located in the brain of the one who has them, then we need to ask: How do qualia inhere in the brain of the one who

has them? Are they properties of brain states, events, neurons, or what? Regardless of which particular aspect of the brain they qualify, we may then ask: Why does observation of the brain not reveal the qualia that the owner of the brain is having? Given the assumption that the brain instantiates the qualia its owner has, should not the brain's qualia be just as observable as the colors instantiated by ordinary physical objects?

This question sounds naive but I submit that it raises a serious problem. The unobservability of qualia in the brain gives rise to the colored-brain problem.[24] Lycan may be right in holding that "no materialist theory of the mind has ever entailed that *watching* the grey cheesy brain of someone who is *having* an intense cyan (or whatever) sensation is qualitatively or in any other way like having that sensation oneself." (Lycan 1990: 111) But, if this is so, it is so only because the materialists went to considerable lengths in their efforts to fortify their views against these naive but embarrassing questions. And the problem of unobservability is not confined to materialism. The property dualist who holds that qualia are nonphysical properties of the brain also owes us an answer to the question why we don't perceive these qualia when we observe the brain of one who is having them.

The colored-brain problem is useful as a structuring device: it lets us classify the various extant accounts of qualia according to how they deal with this problem. The main strategies seem be these. (i) Deny the existence of qualia. Eliminativism is the heroic way out and Dennett has embraced it, as we have already seen. (ii) Acknowledge the existence of qualia but deny that anything physical can instantiate them. The postulation of a nonphysical bearer of qualia—in form of nonphysical sense-data or of nonphysical selves—is characteristic of substance dualist accounts of qualia. (iii) Acknowledge the existence of qualia but relocate them: move them out of the brain and into the world. This might be done, for example, by identifying them with such uncontroversial properties as the colors, shapes, etc., of physical objects. Various versions of the intentional or representational account of qualia rely on this relocation maneuver to come to terms with the problems posed by the unobservability of qualia. Qualia cannot be observed in the brain because they are not there. (iv) Acknowledge the existence of qualia in the brain but recategorize them in such a way that their being in the brain no longer entails their observability. Recategorization forms the heart of the adverbialist account of phenomenal experience. (v) Acknowledge the existence of (unreconstructed) qualia in the brain but explain the crucial notion of *in* in such a way that the existence of qualia *in* the brain no longer entails their observability. Fine tuning the *in* lies at the core of various accounts. Some versions of the identity theory, the dual aspect theory, and of neutral monism belong here. The view I will end up defending also belongs into this camp.[25]

What little I have to say about (i) I have already said. I briefly remarked on the sense-data version of (ii)—as I said, I believe that the case for (ii) is less hopeless than is commonly thought. But I will not attempt to defend (ii) in this context. The discussion of the remaining strategies (iii)-(v) will loom large in the chapters ahead.

2.2.2.9 Background Materialism

Bafflement does not arise in an intellectual vacuum. Qualia are not intrinsically puzzling items; they seem strange only on the background of one's other intellectual commitments. Our particular bafflement is rooted in an unquestioned commitment to materialism. That is, were it not for this background materialism, the existence of qualia might not strike one as strange at all. It is because we started out with the view that all there is are atoms and the void that qualia become problematic. If, instead, one were to choose to view qualia as metaphysically prior, then the existence of matter might strike one as puzzling.

If it is this background materialism that makes qualia problematic, why not simply abandon materialism? There is no short answer to the question why we should be materialists. Jerry Fodor's telling remark that "we're all materialists for much the reason that Churchill gave for being a democrat: the alternatives seem even worse" (Fodor 1992: 5) does a good job of laying bare the main source of our intellectual commitment to materialism. Whether you will be persuaded by this reasoning will, of course, depend on how you determine what makes a metaphysics better or worse. If you happen to be obsessed with the question how qualia fit into the natural order of things, your conviction that materialism is the least bad alternative may start to show signs of erosion. What we need are arguments to prop up our inclination towards materialism. Such arguments as we do find in favor of materialism tend to be variations on the following closely connected themes. Materialism allows us to get rid of repugnant entities and properties; materialism is a philosophy that is in tune with science; and materialism can vastly simplify our cosmological outlook. (See Smart (1963)) These are powerful considerations that we should not brush aside lightly. Nevertheless, they do not provide conclusive reasons for materialism. And some of the most tough minded materialists have clearly seen the ideological component in the commitment to materialism. J.J.C. Smart, for example, tells us that it is "largely a confession of faith" when he says:

> So sensations, states of consciousness, do seem to be the one sort of thing left outside the physicalist picture, and for various reasons I just cannot believe that this can be so. That everything should be explicable in terms of physics…except the occurrence of sensations seems to me to be frankly unbelievable. (Smart 1959: 53-54)

I wholeheartedly applaud this sentiment. One must not give up one's 'naturalistic sobriety' (McGinn's term) at the drop of a hat. I feel the powerful attraction of the naturalistic/materialistic orthodoxy of the current philosophical climate. Materialism seems to be the sanest metaphysics by far. This sermon about sanity and sobriety does, perhaps, deserve to be elaborated just a little bit further. Our overriding intellectual goal is to understand the world. A metaphysics provides the general framework within which one arrives at an understanding of various phenomena that one takes to constitute the world. If the explanatory resources that one's metaphysical framework affords prove too limited to accommodate a purported phenomenon a choice must be made. In some cases it may be wise to sacrifice the phenomenon to reestablish harmony. But if the recalcitrant datum is nonnegotiable, one will have to revise one's metaphysics in the interests of the primary goal of understanding. If the decision to give up materialism is still daunting it is not because of any 'intrinsic self-evidence' of materialism. It is daunting, because of the unparalleled capacity of materialism or materialistic science to help us understand the world. As an explainer materialistic science stands unrivaled. The belief that consciousness will force this giant onto its knees may even seem slightly mad.[26]

Having presented this pitch for materialism, there are, however, two caveats that I would like to register as a warning to the reader. First, I find that I lose my naturalistic sobriety with unsettling ease. The sanity of materialism is obvious to me when, to speak with Hume, I am "merry with my friends." But after I have "indulged in a reverie in my chamber" I am plunged into a "philosophical melancholy and delirium" in which all the obviousness of materialism evaporates without trace. Second, if I were to convince myself that I am incapable of arriving at a materialistically proper account of so obvious a fact as that recorded in our initial datum, then I would consider giving up materialism. That is, materialism is for me a negotiable position.[27]

This tension between qualia and materialism is surely one of the main sources of the qualophobe's phobia and the qualophile's bafflement. Ideally a (noneliminative) account of qualia should resolve this tension. In the final chapter I propose such a reconciliatory account. Whether it deserves the label 'materialism' is an open question.[28] But I hope to show that it preserves the features of materialism that are worth wanting while, at the same time, doing full justice to qualia.

2.3 Two Problems: The Nature-problem and the Having-problem

Given the preceding clarifications the thesis that to be conscious is to have qualia should now be tolerably clear. Two things remain to be done to establish the thesis. First, we need to answer the question: What are qualia? (Call this the 'nature-problem'). Second, we need to answer the question: What is it to have qualia? (Call this the 'having-problem'). Since I try to answer these two questions in the following chapters I will not answer them here. But I do want to make some remarks about how I propose to proceed in tackling the nature-problem and the having-problem.

I have talked at some length about qualia. The upshot is that qualia are the phenomenal properties we meet with every moment of our waking and dreaming lives. What we do not know yet is what these qualia are properties of. So the question what qualia are presents itself as the question: What kinds of objects do qualia qualify? Traditionally, qualia-bearers were taken to be purely Mental or phenomenal objects of some sort. For the purposes of the present discussion, I simply go along with the currently popular view that there are no such purely Mental qualia-bearers. If the qualia-bearers are not nonphysical objects they must be physical objects—actual physical objects or, as has been suggested, nonactual physical objects. Of every proposed qualia-bearer I will ask the following simpleminded question: Can one understand what it is for an object of this kind to exemplify a quale? We will see that not all of the proposed qualia-bearers pass this test.

I have said next to nothing about what it is to have qualia. I have merely employed a large number of metaphorical expressions to stand for the manner in which we have qualia. I have said that we are immersed in a sea of qualia, that we are subject to a ceaseless barrage of qualia, that we are confronted with qualia, that we are inundated with qualia, that we are swamped by qualia, that we meet qualia in experience, that we find ourselves thrown into an ocean of qualia. And I have insisted that it takes no effort, no sophistication, no focused attention, no philosophical process of abstraction or distillation, no cognitive refinement whatsoever to have qualia. Clearly, this is insufficiently precise. These descriptions of what the having of qualia consists in can give us a feel for the kind of relation we are looking for. But they fall far short of specifying this relation between us and our qualia. One constraint on a satisfactory account of the having-problem is imposed by the following consideration. Having qualia is what being conscious consists in. That is, having qualia must account for the fact that it is like something to be us. This suggests a simpleminded question that needs to be put to every proposed account of what the having of a quale consists in. We will have to ask: Can one understand how being related to the quale in

this way makes it be like something to be the subject having the quale? We will see that not all the proposed having-relations pass this test.

Above (Section 2.2.1) I warned against confusing the fact of its being like something for someone with facts concerning one's knowledge of what things are like for one. A similar warning is called for here. What we want to understand is what it is to have a quale. Knowing that one has a quale is quite a different matter and I will not address it here. Our analysandum is the having-relation itself, not one's knowledge that one stands in that relation to something. Here, as everywhere, it is a mistake to substitute the cognition of a phenomenon for the phenomenon itself. This tendency to switch to a more cognitive analysandum goes hand in hand with a tendency to assume that the analysans must contain a cognitive element. The assumption that having a quale must somehow be a matter of knowing (or cognizing) it becomes all but irresistible. Later on (see Chapter Five) I will argue at length that this assumption is unwarranted.

The solutions to the nature-problem and the having-problem are interdependent. The 'location' of the quale will, obviously, have a strong bearing on the question what it takes for us to have the quale. Finding excellent qualia-bearers so 'remote' that it is impossible for us to have those remotely located qualia gets us nowhere. Nor does it do us any good to be given an excellent way of having qualia, if it turns out that we cannot be related to any plausible qualia-bearer in this way. Introspection, for example, will hardly seem like the right way of getting at the quale, if it were to turn out that qualia reside on the surface of external physical objects. Thus solutions to the nature and the having-problem will have to be evaluated as package deals, so to speak.

This interdependence between the nature-problem and the having-problem makes it difficult to discuss them separately. But it is still possible to provide a rough road map of how the following discussion is divided between these two questions. Chapter Five and Chapter Six are primarily addressed to issues surrounding the having-problem. Underlying the discussion in these chapters is a tacit assumption that qualia are, in some sense, internal. This is in keeping with the core idea discussed in these chapters, viz., that having a quale is a matter of introspecting it. The having-problem makes a comeback in Chapter Eight. Here the tacit assumption of the innerness of qualia has been given up and I investigate the question how one is to have qualia that have been jettisoned from the Mind. In Chapter Seven and Chapter Nine the focus is squarely on the nature-problem. The final chapter, Chapter Ten, spends equal time with the nature and the having-problem (hopefully solving both problems in the process).

2.4 Summary

I have tried to explain what I take the term 'consciousness' to stand for. In saying that a subject is conscious I am saying that it is like something to be it. Furthermore, I have attempted to reveal the grounds for my belief that there are qualia and why their existence is obvious, important, and baffling. Though much of the apology for qualia was more of a confession than an argument, I have argued that certain standard objections against qualia need not worry the qualophile. I have urged that we should not begin by assuming that qualia are 'in the mind.' This was the gist of the neutrality thesis. We may, at a later point, be forced to give up the neutrality of qualia. But we should not burden them with a dubious metaphysical status at the outset. The existence of such metaphysically domesticated qualia—the colors, sounds, shapes, smells, tickles, tastes—is blatantly obvious to me. The importance of qualia resides in the fact that it is only by having them that it is like something to be you or me. What could be more important? And the philosophical puzzlement over their existence grows out of an otherwise eminently sensible commitment to a materialist vision of the world: it seems as if qualia (even metaphysically neutral qualia) just should not be here if atoms and the void are it.

We now know what it is that we want to analyze. And hopefully we no longer doubt the existence of the qualia to which the analysis commits us. To say that to be conscious is to have qualia is now no longer an empty phrase. But to show that the thesis is meaningful falls short of showing that it is true. The defense of the thesis will simply consist in spelling out clearly what qualia are and what it is to have them. Possessing an account of consciousness just consists in one's ability to solve the nature-problem and the having-problem.

CHAPTER THREE

On Method: The First-Person Perspective

In the previous chapter I have stated and clarified the thesis I want to defend. I ended by indicating the two problems that a defense of the thesis has to deal with. In the present chapter I present and defend the manner in which I approach these two problems. My thinking about these matters is informed by the first-person point of view. The defense of this approach comes in two parts. First, I must rebut the objections leveled against this manner of proceeding. Second, I must justify the choice of this approach over its competitors. I hope to show that choosing the first-person perspective is a legitimate and fruitful starting point for the study of consciousness. Perhaps this claim will become more plausible when it is seen just how much or how little I include under the heading 'the study of consciousness.' Thus I end this chapter with an assessment of the scope of the project I pursue. I conclude that it is very narrow but very important.

3.1 The Principle of Phenomenological Adequacy

Throughout this discussion I will be guided by a certain policy. The policy is simply to be completely literal minded when it comes to evaluating theses about the location and nature of qualia-bearers and theses about what it is for us to have the qualia they bear. Sometimes there is virtue in sticking to a certain pig-headed literal-mindedness in philosophical discussions. I will reject everything that does not square with what I take to be the phenomenological data. I fear that this strategy may strike some as dogmatic. I do not know how to avoid this impression without betraying the data that seem obvious to me. I hope that the following chapters will show that proceeding in this manner does not cut off discussion prematurely. Paying respect to the phenomenological data does not impose a suffocating stranglehold on philosophical theorizing. It does confine it to some extent, but I consider this an advantage. Lycan's forthright treatment of eliminative materialism nicely illustrates the working of this policy:

> Incidentally, I have been assuming...that people do in fact have mental states that feel certain ways to them—more tendentiously but equivalently, that the

states really do have qualitative or phenomenal characters, however those characters might eventually be explicated. This apparently truistic assumption has actually been challenged [by Churchland and Churchland (1981); Dennett (1982); Dennett (1988); Rey (1983)], who defend the *elimination* of feels at the expense of (anything even remotely approaching) common sense...I would (and do) feel justified in stiff-arming their arguments on Moorean grounds, especially since I am claiming in this book to provide a fully adequate positive account of bona fide phenomenal characters. (Lycan 1987: 63)

While I find myself incapable of accepting Lycan's account of qualia, I applaud his way of dealing with the eliminativist. Not everything goes. The stricture of phenomenological adequacy rules some theories of consciousness out of bounds.

A brief look at one of Armstrong's perceptive methodological comments on his own work on consciousness affords another demonstration of the strictures imposed by the principle of phenomenological adequacy. Commenting on his solution of the puzzles underlying the Sellarsian grain problem Armstrong writes:

> This view [the *gestalt* model] of the secondary qualities is phenomenologically implausible. I once heard Ryle say in his gruff and quasi-military style, at the end of a course of lectures on perception, that there was something 'so perceptual about perception'. He was confessing that this view did not do phenomenological justice to perception. It must also be confessed that the secondary qualities give an overwhelming impression of being ontologically irreducible to the primary qualities. I do not know how to explain this impression. But, in every other respect, the view I arrived at seems to be satisfactory. I am tempted to conclude: 'So much the worse for phenomenology.' (Armstrong 1984c: 28)[29]

'So much the worse for phenomenology' is not a viable option for one who adheres to the principle to phenomenological adequacy. The phenomenology is that which the theory of consciousness is supposed to illuminate. If a theory requires us to disregard the deliverances of phenomenology then it is not the theory I seek. Allegiance to the principle of phenomenal adequacy does not prevent one from entertaining a variety of very different proposals concerning the nature and the having-problem. Most remarkable, perhaps, is the fact that considerations of phenomenal adequacy do not reign in the possibilities canvassed by the neutrality thesis. Qualia may turn out to be properties of inner Mental items; or they may turn out to be properties of the brain or of external physical objects. The phenomenal data do not decide this issue and we will have to follow the qualia around wherever their investigation takes us.

But throughout I will insist on the phenomenological adequacy of the resulting theory of consciousness. That is, I will subject every proposed account of what it is to have qualia to the first-person test of 'trying it on.' I will heed Searle's advice: "Remember, in these discussions, always insist on the first person point of view." (Searle 1980a: 451) The idea underlying this approach is this: "One way to test any theory of the mind is to ask oneself what it would be like if my mind actually worked on the principles that the theory says all minds work on." (Searle 1980b: 417) The application of this strategy to our problem is obvious: if trying on a proposed account of qualia yields the result that having these kinds of qualia in this particular manner is like nothing, then the account must be rejected.

The proposal to proceed in this manner typically elicits an objection that unfolds with the reliability of a reflex. Had we stuck to this policy in the case of physics, physics would still be in its infancy. Therefore, this policy is a bad one. The reply to this objection is equally well entrenched. Different subject matters require different strategies. If we are studying phenomenal consciousness, we cannot hope to get a truer picture of our subject matter by moving away from its surface appearance. For in this case our subject matter *is* the surface appearance. This time we are not trying to understand the reality behind the appearance; this time the appearance itself is the reality we are trying to investigate. I am not recommending the first-person point of view for all inquiries. But I do propose that we should occupy this point of view when our topic is consciousness.

3.2 The Transformation of Consciousness Studies

As the preceding section has made obvious, I undertake this investigation of consciousness from the first-person point of view. The topic of consciousness as well as the first-person perspective have, however, suffered a bad reputation ever since the behaviorist takeover of all thought concerning the mind. But rumor has it that behaviorism is dead. Does that mean that consciousness and its study from the first-person perspective have been rehabilitated? Yes and no. The topic of consciousness has been restored as a legitimate topic. But the comeback of the first-person perspective has yet to happen. I will attempt to show that there is no good reason for this systematic neglect of the first-person perspective in the study of consciousness. Choosing to investigate consciousness from the first-person point of view is an intellectually responsible choice that promises to reveal insights that otherwise would be lost.

In the prebehavioristic days of introspectionist psychology the attempt to present a theory of the mind that was not, at the same time, a theory of conscious-

ness must have seemed like a gratuitous enterprise.[30] The advent of behaviorism thoroughly changed the understanding of what the shape of a theory of the mind ought to be. Consciousness was demoted to a slightly disreputable side issue. But with the demise of behaviorism consciousness moved back into the limelight of research. After a period of relative neglect discussions of consciousness abound in the contemporary literature of the science and the philosophy of mind.

The rebirth of consciousness as a legitimate topic of scientific and philosophical inquiry did not, however, result from a revival of the introspectionist tradition. Consciousness is no longer viewed as that which simply reveals itself from the first-person perspective. It is viewed, instead, as a theoretical construct that the cognitive scientist posits in the light of certain explanatory pressures within his theoretical framework. This change of approach calls for a corresponding change in what can be considered the appropriate method for the investigation of consciousness: the experimental method of cognitive science replaces phenomenology, self-observation, and introspection as the most important tool in the study of consciousness.

Cognitive science is not a monolithic enterprise. The newly-discovered understanding that consciousness is a topic deserving our attention has not led to a corresponding agreement about what *the* problem of consciousness is. But there are at least two unifying themes that clearly sound through the cacophony of theoretical voices: consciousness is understood as a *high level information processing capacity* that has to be studied from the *third-person point of view*. There is no consensus on the precise role that this capacity plays in our mental lives and there is no consensus on the theoretical means that are best suited to account for the appropriately specified capacity. The shared background assumption that the mind is, above all, an information processing device leaves much room for disagreement about how these questions are to be decided. Accordingly, the variations on the information processing theme are numerous. The resulting lack of consensus and the attendant multitude of theories must not, however, lead us to overlook the fact that all the theories of consciousness inspired by cognitive science share at least this other salient feature: they are characterized by their exclusive emphasis of the third-person point of view and the disregard of the first-person point of view that this entails. This bias in favor of the third-person point of view sets them apart from accounts of consciousness that are grounded in the Cartesian first-person perspective.

Thus, the fact that consciousness is, once more, a respectable topic of scientific and philosophical research cannot be construed as the rediscovery of an earlier tradition of inquiry. Theorizing about consciousness is back, but the style of theorizing is different. The role once occupied by the first-person perspective is being usurped by the third-person point of view. Is the first-person point of

view still a viable option, given this powerful trend towards the third-person perspective in consciousness studies? Trends are no arguments. So all will depend on the quality or the reasons that underwrite this trend towards a methodology grounded exclusively in the third-person perspective.

3.3 The Attack on the First-Person Perspective

In the recent discussion Dennett has emerged as one of the most audible and most ingenious advocates of the hegemony of the third-person perspective. I will therefore focus on Dennett's case against the first-person perspective. Dennett's charge has many facets. Here I will pick up only two of these strands. The first strand has two prongs: the friends of the first-person perspective are accused of anti-materialist as well as anti-scientist proclivities. The second strand aims to show that the first-person perspective is obsolete, for Dennett takes himself to have provided a thoroughly scientific (i.e., third-person) method for studying consciousness—the method of 'heterophenomenology'—that has all the advantages and none of the defects of the first-person perspective. A detailed investigation of the method of heterophenomenology (Section 3.3.2) forms the second strand of my discussion of Dennett's case for the third-person perspective. Dennett concludes that the third-person perspective emerges as the only rational choice. I hope to show that the case against the first-person perspective is not yet closed.

3.3.1 The Charge of Obscurantism and the Love of Mystery

Dennett paints the friends of the first-person perspective in dark colors. Obscurantism and the love of mystery inform their choice of starting point. More specifically, the choice of the first-person perspective is an expression of the anti-materialistic and anti-scientific spirit of its adherents. I will, therefore, defend my choice of the first-person point of view by dispelling what I take to be unjustified charges or rumors about this starting point.

When speaking of the choice between the two perspectives, Dennett likes to associate the third-person perspective with materialism and science. He tells us, for example, that he proposes to investigate "just what the mind looks like from the third-person, materialistic perspective of contemporary science" (Dennett 1987b: 7) That is, Dennett makes it look as if points of view came, so to speak, only as parts of larger package deals. With the third-person point of view come materialism and science. Does he want to imply that by choosing the first-person point of view you choose anti-materialism and anti-science? If so, he is wrong.

3.3.1.1 The First-Person Perspective and anti-Materialism

It's possible to combine the first-person perspective with the 'materialistic perspective of contemporary science' because it has been done.[31] Consider three of the most prominent advocates of the first-person perspective in the study of consciousness: Colin McGinn, John Searle, and Thomas Nagel. The ferocity of McGinn's naturalism is unparalleled. Obscurantism and love of mystery (at least of the sort that Dennett fears) have no place in McGinn's world.

> One wants to insist, consciousness cannot *really* be miraculous, some kind of divine parlor trick. It must fit into the natural order of things somehow. Its relation to matter must be intelligible, principled, law-governed. Naturalism about consciousness is not merely an option. It is a condition of understanding. It is a condition of *existing*. (McGinn 1991: 47)

McGinn tells us that the thesis that "nothing that happens is nature is inherently anomalous, God-driven, an abrogation of basic laws" is "an article of metaphysical faith" that "belongs with incredulity over ghosts, telepathy, divine healing and the like." (McGinn 1991: 87) Surely the friends of the 'materialistic perspective of contemporary science' can find no fault with McGinn's way of stomping on those who would turn consciousness into a mystery exempt from the laws of nature.

Of late Searle has turned into an outspoken critic of materialism. So my including him in this lineup may seem puzzling. But if we forget about labels and look at the doctrine Searle clearly qualifies. He holds that mental phenomena are features of the brain, no less biological than, e.g., digestion. (See Searle (1992: 1)) And consciousness is no exception:

> Consciousness is a higher-level or emergent property of the brain in the utterly harmless sense of "higher-level" or "emergent" in which...liquidity is...a higher-level emergent property of H_2O molecules when they are, roughly speaking, rolling around on each other (water). Consciousness is a mental, and therefore physical, property of the brain in the sense in which liquidity is a property of systems of molecules. (Searle 1992: 14)

Whether Searle's 'biological naturalism' deserves to be called materialism need not concern us here. For surely, Searle's account of consciousness is free of the metaphysical malady (dualism) that one allegedly contracts by embracing the first-person perspective.

Nagel's views have been misrepresented with such unfailing regularity that the attempt to style him nondualist may seem quixotic. But a brief look at his pièce de résistance (Nagel 1974) should convince anyone that he is not the foe of materialism he is usually made out to be. After stating his arguments against ex-

tant versions of physicalism, Nagel sums up the results of his discussion as follows:

> What moral should be drawn from these reflections, and what should be done next? It would be a mistake to conclude that physicalism must be false. Nothing is proved by the inadequacy of physicalist hypotheses that assume a faulty objective analysis of mind. (Nagel 1974: 176)

Nagel is well aware that proving the extant versions of physicalism false falls short of proving physicalism false. He acknowledges that we possess "some reasons to believe that sensations are physical processes..." (Nagel 1974: 178) And he closes his paper with an attempt to develop a way of thinking about the subjective character of experiences that would make them "better candidates for objective explanations of a more familiar sort." (Nagel 1974: 179) The point of Nagel's objective phenomenology is to clear the path for a future physicalism: "a phenomenology that is in this sense objective may permit questions about the physical basis of experience to assume a more intelligible form." (Nagel 1974: 179) If we succeed in replacing the 'faulty objective analysis of mind' with a correct objective phenomenology, physicalism may succeed after all. We should view Nagel's cautious ventures into the domain of objective phenomenology as the first steps towards a more promising form of materialistic reduction of consciousness. Though not convinced that such a reductive theory of consciousness must exist, he does not exclude this possibility: "Careful examination will show that no currently available concept of reduction is applicable to [consciousness]. Perhaps a new theoretical form can be devised for the purpose, but such a solution, if it exists, lies in the distant intellectual future." (Nagel 1974: 166) Nagel's work should be understood as a contribution towards this distant goal.

Instead of viewing Nagel (1974) as an anti-materialist tract, we should see it for what it really is: an attempt by a struggling physicalist to show us the direction in which a more satisfying version of physicalism may perhaps be found. It is, therefore, a mistake to categorize Nagel as an anti-materialist friend of mystery. His pronouncements on these issues are more guarded than those of Searle and McGinn—but that is not saying much! Surely, Nagel's stance does not support the thesis that taking the first-person perspective seriously leads to the sort of anti-materialism borne of a love of obscurantism and mystery.

I conclude that the first-person perspective is not guilty of the accusation raised against it. A commitment to this point of view does not entail nor does it arise from some anti-materialistic form of obscurantism.

3.3.1.2 The First-Person Perspective and Anti-Scientism

The upshot of an investigation of consciousness, carried out from the first-person point of view, will not be a scientific theory of consciousness. That is to say, the results of such an investigation will not be couched in the language of the neurosciences, of artificial intelligence, of physics, or of mathematics. To this extent the charge of anti-scientism is justified. If one starts from the first-person point of view one will, instead, arrive at a philosophical account of consciousness. How is one to think of such a nonscientific theory of consciousness? Here is how Panayot Butchvarov characterizes the nature of such an enterprise:

> [Such a] theory may be understood as a philosophical description of the nature of at least some kinds of states of consciousness. Clearly, such a description must be, at least in part phenomenological; it must be defended, at least in part, with phenomenological considerations. For what else can a *philosophical* description of a certain state of consciousness be? Even if a certain state of consciousness is identical with a certain state of a brain, we can know this only if we have an independent conception of the former as well as an independent conception of the latter (to know that the Evening Star is identical with the Morning Star we must first know what we are asserting to be identical with what). And whereas the conception of the latter would be based on anatomical and physiological facts the conception of the former surely can be based only on phenomenological facts. (Butchvarov 1980: 264)

An investigation of consciousness carried out from the first-person point of view will form part of what Brentano called 'descriptive psychology.' This label is not commonly used these days, but the practice itself is alive and well in the analytic tradition. Though not scientific, the result of this kind of inquiry will be a systematic observation-based body of propositions. The observations in question will, admittedly, be ones that have been (and can only be) made from the first-person point of view. But to say that something is observable only from the first-person point of view is not to say that it is not observable. What one observes from this point of view is no less real than what is open to third-person tests.

We are much indebted to Searle for reminding us of the empirical legitimacy of the first-person point of view. Searle holds that two factors conspire to blind us to the relevance of the first-person perspective and the data that it reveals. First, there is the "persistent objectivizing tendency of philosophy and science since the seventeenth century" that inclines us to "regard the third-person objective point of view as preferable to, as somehow more "empirical" than, the first-person, "subjective" point of view." (Searle 1987: 145) Second, we are prone to be misled by

> a systematic ambiguity in the use of the word "empirical," an ambiguity between an ontological sense and an epistemic sense. When people speak of empirical facts, they sometimes mean actual, contingent facts in the world as opposed to, say facts of mathematics or facts of logic. But sometimes when people speak of empirical facts, they mean facts that are testable by third-person means, that is, by "empirical facts" and "empirical methods," they mean facts and methods that are accessible to all competent observers. (Searle 1992: 72)

In combination these two factors lead us to accept the faulty view that the publicly available facts are all the facts there are. We lose sight of the first-person perspective and the empirical facts it reveals.

This shift towards the third-person perspective impoverishes our theorizing about consciousness. A theory of consciousness that limits itself to data sanctioned by the third-person perspective will have to account for a narrower range of phenomena than a theory that also acknowledges data gathered from the first-person perspective. And an exclusive concern with the third-person perspective changes the conditions of adequacy on a theory of consciousness: phenomenological adequacy form the first-person perspective ceases to be an important constraint on such theorizing. The blinkering effect of the predominance of the third-person perspective has not been lost on its partisans. Dennett, for example, has offered the following assessment:

> One of the yawning chasms in the current cognitive theories of the conscious mind is the absence of any clear role for..."the way things feel" or... "qualia"...We thus tend to ignore as unacknowledged implementational noise all the zest, aversion, joy, and anguish that so typically accompany the information-bearing states of our cognitive apparatus. "What could their role be?" we ask ourselves, and, drawing a blank, we divert our attention from them. (Dennett 1990: 100)

Dennett is not one to simply avert his gaze. With great ingenuity and persistence he has tried to show us how the marvelous things that the first-person point of view supposedly reveals either do not exist or can be accounted for from a third-person position—I discuss Dennett's account in Section 3.3.2.

There is, however, no need to choose between the first-person perspective and the third-person perspective. First-person and third-person investigations of consciousness can fruitfully coexist. Instead of simply ignoring or eliminating the data afforded by the first-person perspective, we should attempt to incorporate them into our philosophical account of consciousness. A philosophical account of consciousness that pays no attention to these data will leave us unsatisfied. A completed first-person perspective account of consciousness will then become the target of a third-person perspective theory of consciousness. That is,

a third-person perspective, naturalistic theory of consciousness should be a theory of a philosophically clarified notion of consciousness. First we must have a clear sense of what the problematic feature of consciousness is. Then we can attempt a naturalistic domestication of this feature by reducing it to some scientifically respectable feature of the world.

This way of relating a first-person perspective philosophical account of consciousness with a third-person perspective scientific account of consciousness is what Nagel has in mind when he says that "nothing is proved by the inadequacy of physicalist hypotheses that assume a faulty objective analysis of mind." (Nagel 1974: 176) Elaborating on this theme Nagel says:

> Any reductionist program has to be based on an analysis of what is to be reduced. If the analysis leaves something out, the problem will be falsely posed. It is useless to base the defense of materialism on any analysis of mental phenomena that fails to deal explicitly with their subjective character...Without some idea, therefore, of what the subjective character of experience is, we cannot know what is required of physicalist theory. (Nagel 1974: 167)

The same idea is implied by Butchvarov's claim that "even if a certain state of consciousness is identical with a certain state of a brain, we can know this only if we have an independent conception of the former as well as an independent conception of the latter." (Butchvarov 1980: 264) And Searle expressed the point with characteristic simplicity and bluntness when he observed that many of the recent slew of books on consciousness aren't about consciousness at all. (See Searle (1992: 250))

We can avoid this kind of problem by giving the first-person perspective greater prominence in the study of consciousness. By paying more attention to first-person perspective philosophical analyses of consciousness, we can ensure that our third-person perspective scientific theories of consciousness don't miss their target. These two projects are not in competition with each other. On the contrary, they can coexist and mutually illuminate each other.

The vision of the peaceful and symbiotic coexistence of first-person and third-person inquiries into consciousness may seem a little too cheery and naive to some. Have not the advocates of the first-person point of view traditionally been severe critics of the attempt to shed light on consciousness from an objective third-person point of view? It surely seems that way. It must be conceded that some of the model advocates of the first-person point of view have no such hopes for a symbiotic coexistence of philosophic and scientific accounts of consciousness. Nagel and McGinn have presented related but distinct arguments that seem to indicate an incompatibility of first-person and third-person accounts of consciousness.

But even in the case of the most pugnacious advocates of the first-person point of view the antagonism is not nearly as clear cut as it is often made out to be. First off, it must be stated that neither Nagel nor McGinn have argued that embracing the first-person point view shows that there cannot be a third-person, naturalistic theory of consciousness. What they have argued is that we run into major problems when it comes to understanding how such a theory might be true.

McGinn believes that there can be, and must be a naturalistic, unmysterious account of consciousness. He has, however, also tried to show that it is humanly impossible to grasp such a third-person theory of consciousness. But as far as I can tell McGinn's conclusion cannot be credited to his allegiance to the first-person point of view. Even a cursory review of his extended argument for his verdict of *ignorabimus* (see McGinn (1989)) will reveal that there is no short inference from one's embracing the first-person point of view to the defeatist conclusion. His argument relies on a number of controversial premises that are quite independent of the subjective point of view. Thus, McGinn's case for the impossibility of a humanly available scientific theory of consciousness cannot be tracked down to his embracing the first-person point of view.

Nagel occupies a slightly weaker position. He has argued that we cannot, at present, understand how a naturalistic account of consciousness might possibly be true. But he did not argue that such an account cannot exist. And he has not argued that we will never understand such a theory. He has argued that, at present, such a theory must seem completely mysterious. But so did many theories when they were first proposed. And, as I have tried to show above, Nagel himself has been steadily pushing in the direction of mystery-reduction by working towards an incremental rapprochement of the subjective and objective approaches to consciousness.

Thus, neither Nagel nor McGinn have produced (nor do they take themselves to have produced) anything like an argument to the effect that the choice of the first-person point of view shows the impossibility of a materialistically respectable scientific theory of consciousness. I conclude that one's commitment to science should not scare one away from exploring the first-person point of view. There is nothing intrinsic to this perspective that is anti-scientific in an objectionable way. It is true that an investigation of consciousness from the first-person point of view will not yield a scientific theory of consciousness. But why should it. And it is not true that pursuing a line of inquiry informed by the first-person point of view puts you on an automatic collision course with science. Of course, you get no guarantee that the results arrived at from the first-person point of view will blend in smoothly with scientific theories conceived in the spirit of materialism. But the choice of the first-person point of view leaves the possibil-

ity of symbiotic coexistence open. And this is all one can reasonably demand of a starting point for the philosophical investigation of consciousness.

Those of us who choose to start from the first-person point of view should, therefore, simply refuse to take seriously the suggestion that to opt for this choice is to become a mystery monger of sorts. The fact that some choose to label the less optimistic friends of the first-person perspective with such newly coined (and highly suggestive) technical terms as "the new mysterians" (Flanagan 1991: 312-314) should not intimidate us. By adopting the first-person point of view one does not automatically subscribe to some anti-materialist, anti-scientific form of obscurantism. Making this choice is not something one needs to be ashamed of in these naturalistically inclined times.

3.3.2 The Method of Heterophenomenology

In *Consciousness Explained* Dennett gives a detailed account of the right method for studying consciousness—the method of heterophenomenology. This method is firmly committed to the third-person perspective but designed to capture everything that the friends of the first-person perspective could (legitimately) wish for. As Dennett puts it, heterophenomenology is "a method of phenomenological description that can (in principle) do justice to the most private and ineffable subjective experiences, while never abandoning the methodological scruples of science." (Dennett 1991: 72) When the heterophenomenological analysis of a subject is completed, it yields a "portrayal of exactly *what it is like to be* that subject." (Dennett 1991: 98) What more could the friend of the first-person perspective ask for? If heterophenomenology lives up to its advertisement, then the worry that the third-person perspective leaves out the crucial fact about consciousness, that it leaves out what it is like, seems misplaced. If Dennett is right about the powers of heterophenomenology, then it would seem that my advocacy of the first-person perspective is completely uncalled-for.

I take Dennett's challenge very seriously. It is both promising and threatening to an entrenched devotee of the first-person perspective. It is promising because it does hold out the hope of putting the study of consciousness 'on the secure path of a science.' This is a seductive vision. It is threatening because it creates the unshakable feeling that it is a trick done with mirrors. In what follows I will try to spell out more clearly the nature of this sneaking suspicion. While I would have preferred to present a decisive refutation of heterophenomenology, I can offer nothing of the kind. The best I can do is to uncover some of the crucial presuppositions of the method, contrast them with the assumptions that drive the friend of the first-person perspective, and invite the reader to share my view.

3.3.2.1 What Is Heterophenomenology

What, then, is heterophenomenology? It is "a *neutral* method for investigating and describing phenomenology;" (Dennett 1991: 98) "a neutral way of *describing the data*" (Dennett 1991: 71) for a theory of consciousness. It is the appropriate *descriptive* method for a theory of consciousness inasmuch as it allows one to record only such data as the scientific method permits, i.e., intersubjectively verifiable data. (See Dennett (1991: 71)) It is a *neutral* method of description inasmuch as it does not take sides in "the debates about subjective versus objective approaches to phenomenology, and about the physical or nonphysical reality of phenomenological items." (Dennett 1991: 95) Moreover, heterophenomenology is neutral in the sense of making no prejudgment about whether the subjects it investigates are or aren't conscious. For all the heterophenomenologist knows, the subject under investigation may be a conscious human being, a zombie, or a parrot in a people suit. But despite its scientific neutrality, the heterophenomenological method gives the last word to the subject under investigation—the heterophenomenological subject enjoys constitutive authority over the shape of its heterophenomenological world. (See Dennett (1991: 96))

To see how these three desiderata—scientific objectivity, neutrality, and respect for the heterophenomenological subject—can coherently be realized in a single approach, we will have to lay out the method in more detail. Here, then, is a brief sketch of what the application of the method involves.

> It involves extracting and purifying *texts* from (apparently) speaking *subjects,* and using those texts to generate a theorist's fiction, the subject's *heterophenomenological world.* This fictional world is populated with all the images, events, sounds, smells, hunches, presentiments, and feelings that the subject (apparently) sincerely believes to exist in his or her (or its) stream of consciousness. Maximally extended, it is a neutral portrayal of exactly *what it is like to be* that subject—in the subject's own terms, given the best interpretation we can muster. (Dennett 1991: 98)

The terms 'heterophenomenological world' (or 'heterophenomenology') and 'heterophenomenological object' may seem problematic—Does Dennett really want to say that there are such things? Here is how he explains these notions:

> The heterophenomenology exists—just as uncontroversially as novels and other fictions exist. People undoubtedly do believe they have mental images, pains, perceptual experiences, and all the rest, and *these* facts—the facts about what people believe, and report when they express their beliefs—are phenomena any scientific theory of the mind must account for. We organize our data regarding these phenomena into theorist's fictions, "intentional objects" in heterophenomenological worlds. (Dennett 1991: 98)

In case the nature of these merely intentional inhabitants of heterophenomenological worlds should still strike one as obscure, Dennett explains:

> Heterophenomenological objects are, like centers of gravity or the Equator, *abstracta,* not *concreta.* They are not idle fantasies but hardworking theorist's fictions. Moreover, unlike centers of gravity, the way is left open to trade them in for *concreta* if progress in empirical science warrants it. (Dennett 1991: 96)[32]

The idea that heterophenomenological objects are fictions that can, in principle, be traded in for concreta is crucial. If we manage to identify the theorist's fictions with neurophysiological events, we have explained the nature of conscious objects and showed that they have real existence. If the identification fails, consciousness has been explained away as unreal. Here is how Dennett puts the matter:

> The question of whether items thus portrayed exist as real objects, events, and states in the brain—or in the soul, for that matter—is an empirical matter to investigate. If suitable real candidates are uncovered, we can identify them as the long-sought referents of the subjects terms; if not, we will have to explain why it seems to subjects that these items exist. (Dennett 1991: 98)

Dennett is a little vague about the conditions under which we can go ahead with the identification. But he has no trouble giving a rough sketch of his idea:

> My suggestion, then, is that if we were to find real goings-on in people's brains that had *enough* of the "defining" properties of the items that populate their heterophenomenological worlds, we could reasonably propose that we had discovered what they were *really* talking about—even if they initially resisted the identifications. And if we discovered that the real goings-on bore only a minor resemblance to the heterophenomenological items, we could reasonably declare that people were just mistaken in the beliefs they expressed, in spite of their sincerity. (Dennett 1991: 85)

This, then, is Dennett's idea of how to develop a theory of consciousness that is based on the method of heterophenomenology. Some of the terminology may sound forbidding, but the basic idea is simple enough. We are all heterophenomenologists of sorts. In taking the sounds your interlocutor produces to be the words (the text) 'I have a red-sensation' and in interpreting this text as an assertion that means that she has a red-sensation, you let your interlocutor's words constitute her heterophenomenological world, populated by a lone intentional object (a theorist's fiction)—the red-sensation that your interlocutor takes herself to have. That is, every time you understand someone who tells you how they are, you construct a (partial) heterophenomenological world. What remains to be done is to ask yourself whether what you have been told is true. In raising this

question you do not mean to doubt the sincerity of your informant. What you are asking is merely whether there really is, in her brain, the kind of object that you took her to report, viz., a red-sensation. And this question you answer by doing science. You check whether her brain contains a neural process that can reasonably be called a red-sensation.

This is how heterophenomenology together with neuroscience affords an explanation of consciousness. In the heterophenomenological stage we compile a catalogue of putative objects of consciousness; in the neuroscientific stage we determine whether there are any neural events that are relevantly like the putative conscious events. If yes, then that is what consciousness is. If no, then there is in reality no such thing as consciousness.

3.3.2.2 Changing the Explanandum

The move to heterophenomenology is a disorienting experience for the friend of the first-person perspective. It feels as if everything had been stood on its head. If you begin from the first-person perspective, you take your sensations as the data the existence of which poses the problem of consciousness. The existence of your sensations stands fast for you. The question whether or not they exist or are real isn't even on the horizon.[33] The question you do want to answer is how such phenomena are possible, given everything else you believe about the world. You do not ask: Does consciousness exist? You ask: How is consciousness (the existence of which is beyond doubt) possible?

The heterophenomenologist approaches the issue quite differently. The only hard data available to her are the sounds emitted by her subject. After these raw data have been successively elaborated and interpreted the heterophenomenologist finally arrives at a pale shadow of the autophenomenologist's data: sensations, construed as mere intentional objects, that are posited by the heterophenomenologist in her attempt to make sense of the data she collected.

From this vantage point explaining consciousness is a matter of accounting for objectively observable (and subsequently refined) heterophenomenological data. The sensations posited by the heterophenomenologist are not among these data. The primary explananda for a theory of consciousness grounded in heterophenomenology are "the facts about what people believe, and report when they express their beliefs [about mental images, pains, perceptual experiences, etc.]." (Dennett 1991: 98) The role of the posited sensations is not that of an explanandum but that of a (possible) explanans. If it turns out that these heterophenomenological items are identical with neurophysiological processes, then they are real and their existence accounts for the beliefs that people have in such items. If no such identification is forthcoming, the heterophenomenological items suffer

elimination and all that remains to be explained is "why it seems to subjects that these items exist." (Dennett 1991: 98)

There is a sense in which one might want to say that this sort of heterophenomenological account of consciousness provides or involves an explanation of sensations. But this sense is quite different from the sense in which the explanation of sensations was the goal of the autophenomenological approach. The autophenomenologist asked: How are sensations (the existence of which is beyond doubt) possible? The heterophenomenologist asks: Do sensations exist?

The difference between the explanatory tasks undertaken from the autophenomenological and the heterophenomenological standpoints, respectively, becomes manifest if we consider the consequences of a negative answer to the heterophenomenologist's question. Dennett holds that the correct answer to this question is that sensations do not exist. This negative answer has no tendency to put a damper on the further development of Dennett's heterophenomenological theory of consciousness. All that this negative answer shows is that the task of such a theory is to answer the question "why it seems to subjects that these items exist." (Dennett 1991: 98)

But the discovery, if it is a discovery, that sensations do not exist means the end of a theory of consciousness undertaken from the first-person perspective. For if there are no sensations then there is no deep problem to be explained. The question how sensations are possible is predicated on the existence of sensations. If this presupposition fails the problem goes away. It still remains to be explained why virtually everyone is so deeply mistaken about the nature of their own mental life. But this question raises no deep or disturbing metaphysical problems.

The friend of the first-person perspective will not be able to celebrate this reorientation of the philosophy of consciousness as a great achievement in mystery reduction. You cannot make a mystery go away by simply averting your gaze. And, from the perspective of the autophenomenologist, the decision not to countenance subjective data in your account of consciousness amounts to no more than to averting your gaze from the genuine problem. It is true that sensations do enter the heterophenomenological picture eventually. But they do so in the wrong manner—not as nonnegotiable data, but as disposable theoretical posits.

The switch from the first-person perspective to the third-person perspective literally revolutionizes the study of consciousness. Nothing illustrates this more vividly than a consideration of the change in status that sensations have undergone: once they constituted the glorious mystery of consciousness; now they are in the dock. And Dennett has found them guilty—guilty of not existing.[34] No

wonder that the friends of the first-person perspective are disenchanted with the method of heterophenomenology.

3.3.2.3 Dennett's Official Reply

The first-person perspective inspired animadversions that I have put forward here are nothing new. Dennett has already replied to them. This is what Dennett has to say to all those friends of the first-person perspective who feel that the method of heterophenomenology leaves out the crucial (subjective) facts about consciousness:

> Notice that when you are put in the heterophenomenologist's clutches, *you get the last word.* You get to edit, revise, and disavow *ad lib*...whatever you insist upon is granted constitutive authority to determine what happens in your heterophenomenological world. You're the novelist, and what you say goes. What more could you want? If you want us to *believe* everything you say about your phenomenology, your are asking not just to be taken seriously but to be granted papal infallibility, and that is asking too much. Your are *not* authoritative about what is happening in you, but only about what *seems* to be happening in you, and we are giving you total, dictatorial authority over the account of how it seems to you, about *what it is like to be you.* (Dennett 1991: 96)

This passage is, *prima facie,* puzzling. How can Dennett give me "dictatorial authority" over my account about how it seems to me, about what it is like to be me? In giving me this sort of epistemic authority, he seems to grant me more than even a (cautious) friend of the first-person perspective would want to grant me. For given a sufficiently strong (and not implausible) reading of "dictatorial authority", we can draw the following inference:

1. I believe that I am having a red-sensation.
2. My beliefs about my sensations have dictatorial authority (in some very strong sense).
3. Therefore, I am having a red-sensation.
4. Therefore, a red-sensation exists.

In this way we could establish the reality of sensations, pains, images, etc. Thus the reality of seemings would be established and the case for the first-person perspective would appear strengthened. For it is these seemings that are so precious to the friend of the first-person perspective. And to offer an intelligible account of how these seemings can exist in our world is the aim of a theory of consciousness undertaken from the first-person perspective.

We know that this (i.e., granting the existence of seemings) is not what Dennett had in mind. Dennett's way out is to deny the reality of seemings as distinct from mere judgments:

> You seem to think there's a difference between thinking (judging, deciding, being of the heartfelt opinion that) something seems pink to you and something *really seeming* pink to you. But there is no difference. There is no such phenomenon as really seeming—over and above the phenomenon of judging in one way or another that something is the case. (Dennett 1991: 364)

As Dennett says a few paragraphs further on: there is no actual phenomenology, there only seems to be. (See Dennett (1991: 365-366)) And again: there are no qualia, there only seem to be. (See Dennett (1991: 372))

This denial of real seemings, real phenomenology, real qualia gets Dennett out of the problem that I tried to pose for him. The beliefs over which the heterophenomenologist grants her subject dictatorial authority are not beliefs about real seemings; they are beliefs about yet other beliefs, other judgments. As Dennett puts it in one place: a state of its *seeming-a-certain-way* to you just is a state of your *judging-that-it-is-that-way.* (See Dennett (1991: 133)) From the fact that your belief that you have a red-sensation enjoys dictatorial authority you cannot infer that there is a red-sensation that you have. For having a red-sensation is merely a matter of judging-that-you-have-a-red-sensation. Thus, you have authority only concerning your belief that you are judging-that-you-have-a-red-sensation. But from the fact that you believe (even infallibly believe) that you are judging-that-you-have-a-red-sensation—it does not follow that there is a red-sensation that your are having. For all the epistemic authority vested in you, your lower-level judgment itself might still be false. And even if it were true, you could not get the conclusion that there is a red sensation that you are judging yourself to have.

The friends of the first-person perspective thought that the heterophenomenologist granted them authority over how things seemed to them. Now it turns out that this is not so. The authority granted to the subjects only covers the beliefs about some sort of judgments that they are making. But judgments aren't the stuff of consciousness—at least as far as the friends of the first-person perspective are concerned. It is inevitable that a subject in the heterophenomenologist's clutches should feel cheated once she discovers that this is how she is being interpreted.

Dennett counters by accusing the friends of the first-person perspective of unconscionable whining—for has not the heterophenomenologist given them dictatorial authority about how it seems to them, about what it is like to be them? I reply that the whining is justified after all. For on closer inspection the *seemings* over which you are being granted dictatorial authority turn out not to be the real thing: Dennett substitutes judgments about seemings for real seemings. From the first-person perspective this substitution appears baffling. Having real seemings is what makes it like something to be you, whereas making judgments—be they

about seemings or anything else, for that matter—does not make it like anything to be the one who judges.

3.3.2.4 Phenomenal Realism

Borrowing Michael Tye's term "phenomenal realism" (Tye 1993) for the position that insists on the reality of seemings, I can put the dispute between Dennett and the friends of the first-person perspective as follows. At the root of the disagreement between Dennett and the friends of the first-person perspective lies their disagreement about phenomenal realism. If you accept phenomenal realism, Dennett's heterophenomenological approach to consciousness seems peculiarly oblique, steadfastly aiming right past the center of the issue. On the other hand, if you reject phenomenal realism, then the insistence of the friends of the first-person perspective on there being this further, deeper, objectively uncapturable fact to consciousness appears like mystery mongering.

I think that this debate about the nature of *seeming* is at the core of the debate between Dennett and a friend of the first-person perspective, like myself. It is the seemings themselves that I cherish! I distinguish my seemings from my beliefs in those seemings and from my knowledge of those seemings. And I do not particularly care about the beliefs in these seemings, or even about my knowledge of them. My having the seemings is what makes me different from a zombie. The zombie has all the right (or should we say, the wrong) beliefs about the seemings it does not have. If what you cared for were the beliefs about seemings, then you might as well have yourself zombified without having to fear any significant loss of consciousness. But I would do anything in my power to resist my zombification. For zombification is, in my judgment, just as bad as being dead—it is the cessation of experience, of feeling, of consciousness. Of course my zombified successor has all of the beliefs concerning the seemings, feelings, and experiences it does not have. But these false beliefs are not worth wanting. To have a feeling is to be conscious. To judge that you have a feeling is like nothing. The judgments, without the corresponding feelings to underwrite them, are a pathetic substitute for consciousness.

If you are a heterophenomenologist and begin your investigation of consciousness from the third-person perspective, this speech will not impress you. These so-called data that present themselves from the first-person perspective are, as far as you are concerned, off limits for a responsible investigation of consciousness. The friend of the first-person perspective, on the other hand, will be left with the feeling that the heterophenomenologist persistently misses the essence of what being conscious is all about.

An earlier time slice of Dennett—a kinder, gentler Dennett—used to present a very fair and balanced account of the dispute between the first-person and third-

person perspectives. As Dennett saw it back then, an appeal to argument could not settle the dispute, for any argument for one side would inevitable beg the question against the other. Adoption of either one of these two competing starting points was a matter of rationally permissible *choice*. This is very much how the matter presents itself to me now, after having attempted to defend the first-person perspective against Dennett's more recent challenges. The kinder, gentler Dennett of bygone days never wavered in his commitment to the third-person perspective. The following passage captures this attitude that combines open-mindedness with rationally permissible commitment to the perspective of his choice:

> I propose to see, then, just what the mind looks like from the third-person, materialistic perspective of contemporary science. I bet we can see more and better if we start here, now, than if we try some other tack. This is not just a prejudice of mine—I have shopped around—but the only way I know to convince you I am right is to get on with the project and let the results speak for themselves. (Dennett 1987b: 7)

In the same sort of spirit I want to propose that we set off on the other foot and see what consciousness looks like from the first-person perspective. And the results will have to speak for themselves, just like Dennett says.

I realize that I cannot show the friend of the third-person perspective that the heterophenomenological approach to the study of consciousness is barren. Nevertheless, I strongly believe that the first-person perspective is the much more fertile angle from which to approach consciousness as a philosopher (rather than a scientist). Perhaps your choice of perspective will, in the end, be determined by your conception of philosophy. The conception that is closest to my heart has found its most succinct expression in Hector-Neri Castañeda's words: "Philosophy" he said "is, after all, done ultimately in the first person for the first person." (Castañeda 1979: 2)

3.4 The Scope and Limits of this Project

Let's assume that by now you (i) have a serviceable understanding of qualitative consciousness; (ii) have faith that an analysis of this phenomenon in terms of qualia is not in principle impossible; and (iii) believe that it might be interesting to see this investigation carried out from the first-person point of view. Even if I am allowed to assume all that, you may still demand that I answer some further, strategic, questions before you are willing to read on. You may want to know: What is the relevance of this sort of investigation anyhow? Which of the questions posed by current research projects in the cognitive sciences and the phi-

losophy of mind will this investigation answer? How does it fit into the map of current studies of the mind? These are legitimate questions and I will end this chapter with some sobering reflections about the scope and limits of this project when viewed in the context of the philosophy of mind as a whole.

The project is a narrow one. Just how narrow is it? There are really three questions here. First, How fully do you treat the domain of qualitative consciousness? On this count, the present investigation is quite limited. I will not treat of all the kinds of qualia, not even of all the secondary qualities. As is so often the case, color will dominate the discussion. (In fact it is even worse—nearly all examples will be concerned with the quale afforded by ripe tomatoes!). This is a severe limitation. The suggestion that a similar account could be given of all qualia is, at this point, no more than a pious hope. To meekly wave one's hand while claiming that *mutatis mutandis* the same story could be told across the board would simply be dishonest. The argument for the thesis that this sort of treatment can be generalized to cover the full spectrum of qualitative consciousness has to be made case by case for each qualitative dimension. But this is a project that I do not undertake here.

A second reading of the opening question may be this: How minute a part of the philosophy of mind do you address? It is a small part, indeed. I do not try to offer consciousness as a mark of the mental. I do not attempt to solve the mind-body problem. I do not so much as address the issues of propositional attitudes and mental content. I do not offer a general theory of the mind. I believe that the conclusion one reaches about the nature of qualitative consciousness will have a bearing on all of these issues. But, as far as I am concerned, these are interesting side effects that do not motivate me to undertake this study. That is, I would pursue this investigation, even if I were to convince myself that all other problems in the philosophy of mind could be solved without having an answer to the problem of qualitative consciousness.

However, there is a third and deeper interpretation of our original question. Perhaps the question—Just how narrow is your project?—is not intended as a question about which particular kinds of qualia I do or do not discuss. This whole issue may seem like an in-house dispute among the qualophiles. Perhaps this question does not take me to task for not addressing all the big issues in the philosophy of mind. The question might be intended in a much more profound sense. It might be put like this: How relevant is this kind of inquiry within the wider project of providing stable foundations for psychology? What role can a philosophical analysis of qualitative consciousness play in a naturalistic age in which cognitive science has taken over the questions of the traditional philosophy of mind? Or, somewhat less pompously: Who cares about qualitative consciousness?

This kind of question will present itself forcefully to one who holds with W.V. Quine that philosophy of science is philosophy enough. Given this view of philosophy, one will understand the philosophy of mind as the attempt to clarify the fundamental explanatory notions of scientific psychology. And it takes not great acuity to see that the notion of qualitative consciousness is not one of the central explanatory constructs of the contemporary science of the mind, far from it.[35] Why, then, waste time with qualitative consciousness?

I agree that the philosophy of psychology and the cognitive sciences is a worthwhile enterprise. But I see no reason to go along with the dogma that the philosophy of mind should limit itself to being the philosophy of certain special sciences. I am, therefore, completely unperturbed by the (unlikely) prospect that qualitative consciousness does not and will not play an explanatory role in the sciences of the mind. In general: I can see no reason to accept the maxim that something is a respectable topic for philosophical inquiry only if it plays a central explanatory role in some scientific theory. From the first-person point of view qualitative consciousness presents itself as an absorbing problem that forces itself upon me ineluctably. Consciousness makes it like something to be me. "For it is consciousness, with its power to make the vanishing instant of physical time live on as felt moment of sensation, that makes it *like something to be ourselves*..." (Humphrey 1992: 228) Without it I would be nothing to myself nor would the world be anything to me.[36] I cannot imagine a better reason for singling out an issue for philosophical consideration.

3.5 Summary

In the previous chapter I tried to provide an intuitive understanding of the problematic notions of *consciousness* and *qualia.* In the present chapter I attempted to reveal and defend my methodological commitments. I confessed to being a firm believer in the usefulness of the first-person perspective, and I followed that up with a rebuttal of a number of objections against this starting point. Obscurantism, love of mystery, anti-materialism, and anti-scientism are not part and parcel of the first-person perspective. The heterophenomenologist's (i.e., Dennett's) report of the death of the first-person, autophenomenological approach is greatly exaggerated. Heterophenomenology poses a formidable challenge to the friend of the first-person perspective, but it is one that she can meet to her own satisfaction. I averred that, at bottom, this dispute concerns one's take on phenomenal realism. All this did not amount to a watertight case for the first-person perspective—far from it. But I hope to have shown that choosing this starting point is neither obscurantist, pointless, or irrational. I ended by indicating how narrow

and how important I take this kind of investigation to be. Having dealt with all these (important) preliminaries it is now time to start the investigation proper.

CHAPTER FOUR

Higher-Order Representation and Introspection

4.1 HOR, HOP, and HOT: Terminology

A cluster of new acronyms dots contemporary texts on consciousness: HO, HOT, HOR, HOE, HOP. (See Dretske (1995); Güzeldere (1995)) The acronyms are new but the idea is old: reflection or introspection hold the key to consciousness. British Empiricists, Kantians, post Kantians, and Phenomenologists have held some version of this idea. The central thought animating contemporary versions of this perennially popular approach is this. The mind is hierarchically organized—stratified like a layer cake. Consciousness enters the world when the mind turns back on itself and a higher-level mental state is directed at a lower-level mental state. The lower-level mental state—the target of this inter-level, self-referential maneuver in the mind—is thereby promoted into a conscious state. This sketch of an idea has been developed in different ways giving rise to various versions of what I will call the 'hierarchical' or 'higher-order' or 'higher-order representation' theories of consciousness—HO or HOR theories for short. The current debate is dominated by two versions of the HOR model. On the one side we find those who hold that the higher-order mental state is a thought. This is the HOT theory of consciousness. On the other side we find those who hold that the higher-order mental state is a perception or experience. This is the HOP or HOE theory of consciousness. I will use the label HOR to refer in a generic way to any version of a hierarchical theory of consciousness.[37]

With this terminology in hand I can briefly outline the plan of this chapter. I want to defend three claims. First, I argue that both versions of HOR can be understood as theories of qualitative consciousness, *pace* the apparent disclaimers of their proponents. Second, I present an analysis of why it is so difficult to pin down the intended explanandum of HOR. The third claim I defend concerns the relationship of HOP and HOT and the traditional introspection-based theories of consciousness (or 'introspectionism'). I argue that HOP and HOT have much in common and that this shared doctrine is identical with introspectionism. All of

this material is preparatory for the critique of introspectionism that will take up Chapter Five and Chapter Six and a good part of Chapter Eight. What I want to persuade you of eventually is that neither the thinking of higher-order thoughts (HOT) nor the having of higher-order perceptions (HOP) can serve to switch on the inner light. The present chapter sets the stage for this critique of introspectionism.

4.2 Is HOR A Theory of Qualitative Consciousness?

The discussion concerning the intended explanandum of HOR comes in three parts. I begin by presenting textual evidence that seems to make crystal clear that HOR is *not* intended as a theory of qualitative consciousness. I then consider the case of David Armstrong's absentminded driver. Brief reflection on the nature of the mental state of Armstrong's driver will appear to afford another piece of evidence for the claim that HOR is *not* intended as a theory of qualitative consciousness. Prolonged reflection on the case of the absentminded driver does, however, undermine this verdict. Finally, I present my reasons for thinking that HOR ought to be viewed as a theory of qualitative consciousness. By the end of this discussion I hope to have established that HOR is accountable for qualitative consciousness. Only then will we be in a position to criticize HOR for not offering a satisfactory account of what it is like to, for example, see a ripe tomato.

4.2.1 The Independence of HOR and Qualitative Consciousness

The attempt to challenge HOR with the problem of qualitative consciousness may seem misguided. For the leading defenders of the HOT and the HOP incarnations of HOR appear to be agreed in claiming that their respective theories are *not* intended as theories of qualitative consciousness. If this is so, my point against HOR is quite pointless, as it were.

Here I will focus on the work of David Rosenthal and William Lycan whose names have come to be identified respectively with the HOT and HOP approaches to consciousness. Lycan has put the case as clearly as one could wish for: "Contrary to some philosopher's misreadings of the "internal monitoring" literature, monitoring consciousness or introspective awareness has virtually never been claimed to explicate phenomenal character…" (Lycan 1995c: 263, fn. 1) And in a recent exposition of his version of HOP he writes:

> I am not here addressing issues of qualia or phenomenal character…There may be Inner Sense theorists [Lycan's term for HOP] who believe that their views solve problems of qualia; I make no such claim, for I think qualia

> problems and the nature of conscious awareness are mutually independent and indeed have little to do with each other. (Lycan 1995b: 2)

Since I understand consciousness as the having of qualia, it appears that we cannot do business. Consciousness, as Lycan understands it, has nothing to do with problems about qualia.

Rosenthal has written an aptly titled paper called "The Independence of Consciousness and Sensory Quality." (Rosenthal 1991) There he argues that

> the properties of being conscious and having sensory quality are independent of one another, and a satisfactory account of each property requires us to investigate them separately. (Rosenthal 1991: 16)

The proper subject for the HOT theory to investigate is, according to Rosenthal, consciousness and not sensory quality, or qualia. Again it looks as if HOR, this time in its HOT version, has nothing to say about the qualia related issues that I am interested in. It would seem, then, that the most prominent expositors of the two versions of HOR are at one in unequivocally rejecting the challenge with which I want to confront them. As they see it, qualia, phenomenal character, sensory quality and all that goes into qualitative consciousness are not the explanandum of HOR. So my point that HOR does not account for qualitative consciousness seems to count for nothing.

4.2.2 Armstrong's Absentminded Driver and the Explanandum of HOR

On to another piece of evidence in support of the view that HOR is not intended as a theory of qualitative consciousness. David Armstrong is the author of the notorious long-distance driver example:

> After driving for longer periods of time, particularly at night, it is possible to "come to" and realize that for some time past one has been driving without being aware of what one has been doing. The coming-to is an alarming experience. It is natural to describe what went on before one came to by saying that during that time one lacked consciousness. Yet it seems clear that...there was perception. That is to say, there was...perceptual consciousness. If there is an inclination to doubt this, then consider the extraordinary sophistication of the activities successfully undertaken during the period of "unconsciousness." (Armstrong 1978: 59)

What are we to think of the mental state of the driver? For our HOR theorists (as well as for Armstrong, himself an early champion of HOP) the situation is clear. First, the driver does have perceptual, more specifically, visual consciousness of the road, the signs, the lights, and the traffic he reacted to. But this form of consciousness is *not* the explanandum of HOR theories of consciousness. Second,

the driver lacks the form of consciousness that HOR tries to illuminate. He has no conscious awareness of the sights and sounds he perceives while in his reverie. (Compare Lycan (1995a: 250); Rosenthal (fc.: Section V)) I, however, am inclined to say that Armstrong's driver does enjoy what I call qualitative consciousness during his absentminded spell. It would seem, then, that the sort of consciousness I am interested in is not the sort of consciousness that HOR theories of consciousness try to account for. So we appear to have one more reason to think that it is a mistake to fault HOR for not accounting for qualitative consciousness. HOR theories were not designed for that purpose. Again we have to conclude that the fact that HOR does not account for qualitative consciousness counts for nothing—or so it would seem.

4.2.2.1 Reflections on Armstrong's Driver

But the lesson Armstrong's driver example teaches is not that simple. Further reflection on the case raises doubts as to what sort of consciousness the absentminded driver does or does not have. Two questions need to be asked. First, is it clear that the sort of consciousness the driver has is not the sort that HOR attempts to capture? The attempt to answer this question naturally leads to the second question: Is it clear what is involved in the perceptual consciousness attributed to the driver? As we will soon see, it is surprisingly difficult to answer these questions within the hierarchical framework. The discussion of these questions will reveal certain unexpected limitations of HOR, and it will thoroughly undermine the notion that Armstrong's driver case can be used to clearly distinguish HOR's intended explanandum from qualitative consciousness.

Let's begin with the first question. Two facts are obvious. The driver is better off than a blindsighted person. And the driver lacks what we might call 'fully introspective consciousness'—the sort you enjoy when you purposefully and attentively focus on your experience. Thus, on the 'spectrum of consciousness' the driver's consciousness must be situated above that of the blindsighted person and below that of the fully introspectively conscious person.[38] This intermediate kind of consciousness will come into sharper focus after the extremes have been pinned down. On the high end of the spectrum we have fully introspective consciousness. To enjoy fully introspective consciousness one must, according to Rosenthal, have a third-order state that makes conscious the second-order state that makes conscious the target mental state. (See Rosenthal fc.: Section VI) Lycan requires that the internal attention mechanism that makes a mental state conscious be "deliberately mobilized by their owners" rather than doing its work by merely functioning automatically. Only then do we deal with genuine introspecting and the corresponding high-grade, fully introspective consciousness. (See Lycan (1995b: 11))

Turning to the low end of the spectrum of consciousness we must acknowledge the sobering fact that nothing about blindsight is clear or obvious. I discuss it at some length in the Chapter Six (see especially Sections 6.4.1-6.4.9). In particular, I focus on John Pollock's HOR-based characterization of the phenomenon:[39]

> Blindsight is an amazing phenomenon, and most people's initial reaction is disbelief. But sense can be made of the phenomenon in terms of the distinction between perceptual sensors and introspective sensors. The very fact of discrimination indicates that the patient's perceptual sensors are still operating. But he is unaware that they are operating, and what that indicates is that his introspective sensors are not detecting the operation of his perceptual sensors. For this reason he has visual sensations, but they have no *qualitative feel.* (Pollock 1989: 16)[40]

The blindsighted person, then, is one whose mental states are neither introspected (HOP) nor thought about (HOT). The mental states of the fully introspectively conscious person are accompanied by a deliberately guided attention (HOP) or a couple of stratified thoughts. (HOT).

Given these HOR-based characterizations of blindsight and fully introspective consciousness, the hierarchical framework naturally suggests a characterization of the intermediate degree of consciousness that Armstrong's driver enjoys. The HOP theorist will have to say that the driver's perceptual states will be monitored automatically by her attention mechanisms (rather than deliberately, as in the case of fully reflective consciousness, or not at all, as in the case of blindsight). The HOT theorist will have to say that her perceptual states are accompanied by a single higher-order thought (rather than by two such thoughts, as in the case of fully introspective consciousness, or no such thought, as in the case of blindsight).

But this is not a happy result for the HOR theorists. For it commits them to the view that Armstrong's driver—everybody's favorite *un*conscious vehicle operator—is conscious in precisely the sense that the HOR account tries to capture. The HOR theorist will protest that this will not do as an answer to the first question. To him the question—Is it clear that the driver's consciousness is not of the sort that HOR attempts to capture?—is merely rhetorical. For to him it is obvious that the driver's consciousness differs from the sort of consciousness that HOR captures. I.e., to him the answer is clear and it is Yes. But our considerations have not borne out this intuition. We too arrived at a clear answer, but it is No rather than Yes.

So much, then, for an answer to the first question. The discussion that follows takes off from reflections on this answer but it will lead us straight into a discussion of the second question concerning the precise nature of the driver's

perceptual consciousness. The friends of HOR will not let this answer stand, I am sure. But no matter how much success they have in repairing the damage, the point has been made that the lesson taught by Armstrong's driver example is not so simple. And as we pursue the ensuing moves in the unfolding dispute the shadow cast over the driver example deepens.

I foresee two reactions against the argument presented above. First, one might hold that it is wrong that the driver's consciousness is situated between the extremes of blindsight and fully introspective consciousness. Nobody, I trust, will argue that the driver enjoys fully introspective consciousness. So the objection amounts to the claim that it is wrong that the driver's consciousness is of a higher grade than that of the blindsighted person. Second, one might hold that blindsight was mischaracterized.

The first reply would be an elegant way out if it were believable. But it isn't. The mental state of Armstrong's driver is quite common—we all experience it much of the time. How, for example, did you experience the printed characters you just read? Nothing comes to mind. Now, after I asked the question, you are fully introspectively conscious of your experience of the print. But for the last hour you saw and reacted to the print, even though the experience of the print (not the message, I hope) passed through your mind without leaving a trace. However, you would be quite mistaken to congratulate yourself on your bravura blindsight performance. For you did have what the blindsighted person lacks. As both Lycan (fc.: 29) and Rosenthal (fc.: Section V) acknowledge, you (and the long-distance driver) were perceptually conscious. Lycan, for example, provides the following characterization of the mental state of Armstrong's driver just after he stopped for a red light:

> Presumably the light looked red rather than green to him; that is the only reason he would have stopped. So, in our present strict sense of the term, he was presented with a red quale; a subregion of his visual field had redness as its phenomenal or qualitative character. But the driver was not aware of any such thing; it was an un- or subconscious perceiving, entirely unintrospected. (Lycan 1995a: 250)

This is a characterization of what it is like to be absentminded. It is not a characterization of what it is like to be blindsighted. All parties to the dispute seem to agree on this.[41] So we can discard the first reply.

The second reply charges that the apparent problem for HOR arose because blindsight was mischaracterized. The difficulty stemmed from the fact that the two points marked by the notions of blindsight and fully introspective consciousness lay too close together on the spectrum of consciousness. Thus, the space that the driver's intermediate sort of consciousness could occupy was narrowed down to a point—the wrong point, as it were. For this point is already

taken by the sort of consciousness that HOR theories of consciousness address. We need a revised, more primitive notion of blindsight. This simplification will increase the distance between the redesigned notion of blindsight and the notion of fully introspective consciousness. By thus stretching the spectrum of consciousness we enable it to accommodate two intermediate notions of consciousness between its extremes: the sort of consciousness that HOR theories capture and the simpler sort of perceptual consciousness enjoyed by Armstrong's driver.

Here is another way of making the point that I underestimated the width of the spectrum of consciousness by mislocating its lower end. Lycan's characterization of the driver's consciousness (see p. 64) suggests that we need to seek a simplified notion of blindsight. For the perceptual consciousness that Lycan attributes to the driver corresponds precisely to Pollock's notion of blindsight—the notion with which we have been working so far (see p. 63). I.e., perceptual consciousness, so understood, now resides at the lower limit of the spectrum of consciousness. But that is wrong because the state of consciousness enjoyed by a blindsighted person is more primitive than that of the driver. It follows that the HOR theorists will have constructed an appropriately simplified notion of blindsight that can now cap the spectrum at its lower end.

It is helpful to view the HOR theorist's objection as an instance of the strategy stated above, viz., to attempt to get the driver's consciousness into sharper focus by pinning down the extremes between which this sort of consciousness is located. The present objection holds that my image of the driver's consciousness became all unfocussed and blurred because I did a poor job of pinning down the lower end of the spectrum along which the perceptual consciousness of Armstrong's absentminded driver is positioned. Move the lower end to the left and the driver's consciousness will naturally fall into place.

What we need, then, is a more primitive notion of blindsight. Can we have it? One problem I foresee arises from the criterion of perceptual consciousness that all of our HOR theorists have been employing. Armstrong argues that we can ascribe perceptual consciousness to the driver on the basis of "the extraordinary sophistication of the activities successfully undertaken during the period of "unconsciousness."" (Armstrong 1978: 59) Lycan concurs when he emphasizes the connection between perceptual consciousness and actions on the part of the driver: "He must have perceived the bends in the road, the road signs, the stop lights, and so on. Suppose he did in fact stop at a red light. Presumably the light looked red rather than green to him; that is the only reason he would have stopped." (Lycan 1995a: 250) And Rosenthal makes the same point when he writes: "To make one's way when driving or walking somewhere, one must be visually conscious of many obstacles. But one may well be wholly unaware that one is conscious of those obstacles if, for example, one is immersed in conver-

sation." (Rosenthal fc.: Section V) At a minimum these passages suggest a sufficient condition for perceptual consciousness: if you can exploit certain forms of information about your environment to guide your behavior, then you are perceptually conscious.[42]

The problem we now encounter is this: when applied to blindsighted persons, this criterion yields the result that blindsighted persons are perceptually conscious. A point not lost on Pollock—himself a HOR theorist—who based his problematic characterization of blindsight on this very insight: "The very fact of discrimination indicates that the [blindsighted] patient's perceptual sensors are still operating." (Pollock 1989: 16) Blindsighted individuals do display the sorts of behaviors upon which Armstrong, Lycan, and Rosenthal base ascriptions of perceptual consciousness. Hence we get the result that, whatever else may be true of blindsighted persons, it will also be true that their state of consciousness is of the perceptual variety. As long as we rely on the criterion of perceptual consciousness preferred by the HOR theorists, we will not be able to mark an interesting difference between the conscious states of absentminded people and blindsighted people. And that is not true.

We can approach the same result from a slightly different angle. The focus here is not on the criterion of perceptual consciousness but on the machinery that drives the capacities of blindsighted subjects. It must be acknowledged that the actual behavior of blindsighted patients is less sophisticated than the behaviors we display while acting on 'autopilot.'[43] This opens up the possibility to revise the criterion of perceptual consciousness in such a way as to take advantage of this performance difference. Perhaps that can be done in such a way as to not seem entirely *ad hoc*. But even so a nagging question remains. Blindsighted subjects do react sensitively to certain features of their environment. Since they don't do it by magic, they must do it by discriminating between their internal states that are informationally correlated with the relevant states of their environment. Now Rosenthal, for example, has defined the properties upon which such discriminatory ability turns as sensory qualities or qualia (see, e.g., Rosenthal (1991: 26); Rosenthal (fc.: Section II)). It follows that anybody capable of such discriminations has qualia. And, given Lycan's explication of perceptual consciousness (see p. 64 above), it follows that everybody capable of such discrimination is perceptually conscious. In the light of this argument, it appears that the attempt to exploit subtle behavioral differences between blindsighted and absentminded persons will not suffice to drive an interesting wedge between them inasmuch as their perceptual consciousness is concerned.

These considerations show that there is no simple way for the HOR theorist to draw a clear distinction between the sorts of consciousness attending the mental states of blindsighted and absentminded persons. HOR makes it difficult

to specify a series of 'degrees of consciousness' that covers the range from blindsight, at its minimum, to fully introspective consciousness, at its maximum. This may not be immediately obvious when we look at the different segments of the spectrum in isolation. But it becomes quite striking when one tries to lay out the whole spectrum at once. Thus, hierarchical accounts of consciousness are still faced with the problem of providing an account of the mental states of Armstrong's long-distance driver that is intuitively sound and theoretically acceptable. This difficulty—i.e., the difficulty of providing a HOR based answer to the second question (see p. 62)—is connected with the problems we encountered when addressing the first question (see p. 62). As long as it is not clear what the nature of the driver's consciousness is it will remain unclear whether it is of the sort that HOR attempts to explain.

The discussion of Armstrong's absentminded driver has proved to be of considerable intrinsic interest. We have seen that HOR may be too limited to encompass the various degrees of consciousness required to make sense of the relevant differences between blindsighted, absentminded, and fully aware drivers. I take this to be a telling example of the difficulties one encounters when one bases one's theorizing about consciousness on the unduly restricting principles of the hierarchical framework. At this point we should, however, remind ourselves why we embarked on this digression. It seemed as if consideration of Armstrong's driver could provide a strong reason for holding that HOR does not address itself to qualitative consciousness. We can now conclude that, contrary to first impressions, this is not so.

I have made much of the difficulty of distinguishing between the sorts of consciousness attending the mental states of blindsighted and absentminded persons. Working within the confines of HOR this seemed like a genuine problem. But it is also prone to seem like an artifact of this theoretical approach. For the problem vanishes when considered from without the framework of HOR. The difference between the sorts of consciousness accompanying the mental states of blindsighted and absentminded person is profound and easily characterized. It is like nothing to be a blindsighted person making visual discriminations. But it is like something to see, e.g., a red light, even if one does so absentmindedly. Or, as I would like to put it, absentminded people do, but blindsighted people do not, *have* qualia. The HOR theorists cannot avail themselves of this simple way of drawing the distinction. For it involves an appeal to qualia and 'qualia problems' are off limits, or so we have been told (see Section 4.2.1).

Thus, we are back to the question whether HOR should be considered a theory of qualitative consciousness. We saw the passages in which the leading figures of the hierarchical school of thought seem to deny any responsibility for qualitative consciousness. I want to now present the case for holding that HOR

does aspire to address the problem of phenomenal consciousness. Then I will briefly comment on the question whether broadening the scope of HOR makes it easier to distinguish the blindsighted from the absentminded and answer in the negative. The discussion closes with an analysis of the difficulty involved in determining whether the hierarchical theory of consciousness is or is not to be understood as a theory of qualitative consciousness.

4.2.3 HOR and What It's Like

Having presented the textual support for the thesis that HOR has nothing to do with qualitative consciousness, I spent considerable time undermining what might have seemed like a strong argument for this view. It has become obvious that appeals to Armstrong's driver cannot settle the question of the intended explanandum of HOR. But the textual evidence against the thesis that HOR is a theory of qualitative consciousness still stands. I now present the reasons for thinking that HOR does address itself to qualitative consciousness. After all is said and done it will, I hope, become clear that the purported textual evidence against my thesis does not show what it so clearly seems to show.

After Lycan's explicit denials that his inner-sense version of HOR deals with phenomenal character or qualia problems of any sort it comes as some surprise when we read: "the Inner Sense account affords the best solution I know to the problem of subjectivity and "knowing what it's like," raised by B.A. Farell, Thomas Nagel and Frank Jackson." (Lycan 1995b: 3) Here is the magic phrase, associated with all the right names too. So, contrary to first impressions, HOR does address itself to the sort of consciousness that is captured by the 'what it's like' locution.

Nor is Lycan alone in holding this view. For Rosenthal follows right along, asserting the HOT theory's ability to shed light into the darkness that is the inner illumination.[44] Having a thought about a sensorily qualitied state is sufficient for one to be qualitatively conscious: "when states with sensory qualities are conscious, there will be something it is like to be in those states…" (Rosenthal 1991: 22) And it is also necessary: "there is no such thing as what it is like to have these sensations unless the sensation is conscious." (Rosenthal fc.: Section II) In a word, the HOT story is the *complete* story about qualitative consciousness. We have qualitative consciousness just in case a state with sensory qualities is accompanied by a higher-order thought. I believe that this is the view that Rosenthal means to express when he writes: "What it is like to have the pain…is simply what it is like to be conscious of having that pain." (Rosenthal 1991: 19)

Both authors have much to say about the nature of the higher-order representations which, together with our qualitative states, give birth to qualitative con-

sciousness by making *felt* the intrinsically unfelt qualitative properties of our mental states. (See Lycan (1987); Lycan (1990); Rosenthal (1986); Rosenthal (1993); Rosenthal (fc.)) The crucial question whether representation (of any order) can perform the task that Lycan, Rosenthal, and many others assign to it will be taken up in Chapter Six and especially Chapter Eight. Setting aside this crucial question for the moment, we can still draw the following limited but important conclusion from the preceding discussion of the HOR theories of consciousness: qualitative consciousness does fall within the intended domain of HOR theories of consciousness. Thus, if HOR theories of consciousness do not manage to explain qualitative consciousness they are, to that extent, failures.

This humble conclusion will not come as a surprise to HOR theorists. But the HOR camp's loud protestations that they have no truck with phenomenal character or sensory quality, etc., have caused considerable confusion in the anti-hierarchical party. These denials are apt to create the impression that HOR has been immunized against the most serious challenge with which the foes of HOR can confront it, viz., the challenge of accounting for qualitative consciousness. But in the eyes of HOR's opponents this immunity was purchased at the price of insignificance: a theory of consciousness that does not address qualitative consciousness does not merit much attention. This impression is false. HOR does address itself to qualitative consciousness, it does merit our attention and it can be held responsible for explaining why, for example, it is like something for one to see a ripe tomato.

Before presenting an analysis of this somewhat bewildering situation concerning the intended explanandum of HOR, I want to offer one last comment on the difficulty that HOR encountered in characterizing the conscious states of Armstrong's driver. The material presented in this section shows that HOR is a theory of qualitative consciousness. I.e., HOR claims to tell us "What What It's Like Is Really Like." (Tye 1995b) If this is so, one is led to expect that HOR might, after all, possess the resources with which to capture the intuitively obvious distinction between the consciousness of the blindsighted and the absent-minded person. But this is not so because, according to HOR, the what-it's-like dimension appears on stage only after the appropriate second-order representations have been tokened. And a crucial feature of the driver example is, of course, that she does not entertain the requisite higher-level representations. So HOR would have it that it isn't like anything to be conscious after the fashion of Armstrong's driver. This means, in effect, that the inclusion of the what-it's-like dimension in HOR's prospect does not change what has already been said on the issue of the driver's consciousness. The driver has unconscious qualia that do not have a feel attached to them. Thus we are led right back to the worry that the mental state of the absentminded driver (as seen through the eyes of the HOR

theorist) becomes indistinguishable from that of the blindsighted person. Expanding HOR into a theory of qualitative consciousness does not help to overcome this puzzle.

4.3 Why Is HOR's Explanandum Elusive?

Using the terminology introduced in Section 2.3 the debate about the explanandum of HOR can be stated succinctly. HOR does not address the nature-problem. But HOR does address the having-problem. When Lycan and Rosenthal (qua HOR theorists) renounce all responsibility for phenomenal character and sensory quality they are saying that HOR is not a theory of the nature of qualia. I.e., HOR does not say what sorts of properties qualia are, what their metaphysical or categorial status is, how they are instantiated, whether they are reducible, and so on. But HOR does address the having-problem. HOR tells us how you have to be related to a quale—whatever its nature may be—so that having the quale in this way results in its being like something for you. Take, for example, the red quale involved in seeing a ripe tomato. HOR tells us that there is a mental state (or an intentional object of a mental state) that instantiates sensory redness plus a higher-order representation directed at that state. The combination of these two factors makes it be like that for you to see a ripe tomato.

The distinction seems simple enough. So simple in fact that one may wonder how anybody might have missed it. I believe that there are deeper (if not good) reasons for the impression that the explanandum of HOR is elusive. I argue that the difficulty to pin down the intended explanandum of HOR can be traced back to a significant ambiguity in the notion of a quale.

4.3.1 Qualia: Celluloid vs. the Silver Screen

The use of the term 'quale' (pl. 'qualia') (and all of the roughly synonyms listed in Section 2.2.2.3) has undergone a disorienting transformation. This term was introduced to stand for the features of experience that engulf you when you are alert and enjoying the experience, as it were. The focus here is on the core cases of full-blooded, lived experience. The extremes—trained phenomenologists and obsessive introspectors on the one end, blindsighted patients and completely spaced-out motorists on the other end—we exclude in the interest of fixing the idea clearly. Qualia are the features of these paradigmatic experiences; they are the colors, sounds, tastes, smells, etc., that make up what it is like to be you at that moment.

But a new usage of the term 'quale' has established itself. Now one can say that someone is in a state exemplifying intense pain qualia but that she does not

feel a thing. Or one might openly wonder whether a blindsighted person might have perfectly intact visual qualia. At one time such speculations would have been greeted with the incredulous stare that we reserve for those who utter absurdities. But contemporary ears are not offended by such statements. They hear no hint of the jarring dissonance that the traditionalist perceives here. Qualia are no longer understood as designating features or aspects of the glorious end-state of a fully consummated experience. Now that term stands for features of some inner state of yours (Rosenthal) or some intentional object of yours (Lycan) that may or may not play a role in what it is like to be you at the time. Only if the quale in question is subject to higher-order representation (of the right sort) will it contribute to what it is like to be you. The quale, all by itself, is not a constitutive element of lived experience. Only by being represented is it empowered to do that. Unrepresented qualia pass through the mind unfelt.

A motion-picture analogy suggests itself. Originally, qualia were features of the brilliant spectacle unfolding on the theater's screen. Now they are features of a strip of celluloid (i.e., the film), the videotape, the laser disc. Features of the show turned into features of the shown medium. Unlike the show's features, the features of the medium need not be shown to exist. An unprojected film is none the worse for that. But an unprojected film does make for a poor show. Exploiting this analogy, we can dub the two contrasting views of qualia the *silver-screen* and the *celluloid* view of qualia. Celluloid qualia can be exemplified without being represented. But an unrepresented celluloid quale of yours does not contribute to what it is like to be you. In order to do that you must represent it with higher-order thought or perception. You silver-screen qualia, on the other hand, need not be represented in order to form a part of your qualitative consciousness. But unlike a celluloid quale, a silver-screen quale cannot be exemplified in the darkness behind the stage of consciousness.

Putting matters in these terms, we can say that (i) HOR works with celluloid qualia; (ii) HOR is not a theory of the specifics of the medium (the film, the tape, or the disc) that exemplifies these qualia; and (iii) HOR does provide a theory that tells you how to transform the information in the medium into the brilliant picture on the screen.

The contrast between the two conceptions of qualia is profound. If one starts with the traditional idea of silver-screen qualia the toned down celluloid qualia do not seem like qualia at all. They are merely cogs in the complex machinery that brings about the instantiation of real (i.e., silver-screen) qualia as its result. Only when 'the light goes on' do qualia make their appearance. Anything that can be fully present in the darkness behind the curtain is not the real thing. The champion of celluloid qualia may be willing to admit that his qualia are less glamorous than the silver-screen qualia they supplant. But he will take pride in the fact that

they, unlike their mythical forebears, exist. That is, they can actually be instantiated by objects in our world—a privilege that the old silver-screen qualia never enjoyed. Thus it may seem that switching to celluloid qualia has an enormous advantage. Celluloid qualia are naturalistically well-behaved properties that brains and tomatoes can instantiate with equal ease. For celluloid qualia the nature-problem simply does not arise.

But the gains on the side of the nature-problem are offset by losses on the side of the having-problem. Two problems face the friend of celluloid qualia. First, working with celluloid qualia makes it hard to account for the striking features of qualitative consciousness. The distance between the very plain celluloid qualia and the extravagant features of lived experience is large and an enormous burden is placed on the machinery that transforms celluloid qualia into conscious experience. On everybody's reckoning the machinery that is supposed to accomplish this task is the mechanism of representation. Whether representation is up to this task is the topic of Chapter Six and especially Chapter Eight.

The second problem is less easily stated. Consider again the person who instantiated pain qualia of which he is not aware. At the moment in which she becomes aware of the hitherto unnoticed qualia her subjective experience takes a terrible turn for the worse. A natural way to describe this situation would be to say that this change consists in the victim's coming to instantiate a whole new set of properties—properties quite distinct from the celluloid qualia that she exemplified all along. And these new properties that make up what it is like to experience the pain look very much like the traditional silver-screen qualia. But the convert to celluloid qualia cannot put it that way—the whole point of adopting celluloid qualia is to rid oneself of the mysterious silver-screen qualia. Thus the second problem facing the champion of celluloid qualia comes to this: the factor constituting the salient augmentation of our experience that sets in at the moment we notice the qualia we instantiated throughout is not easily captured in a framework limited to celluloid qualia. A number of philosophers have attempted to solve this problem with an appeal to illusion. The person noticing her pain quale does not thereby acquaint herself with a wholly new property. In noticing the quale she grasps the celluloid quale but in a less then perfect manner that contains a substantive admixture of illusory components. Thus, the enriching addition to the experience cannot be traced to the instantiation of a new and mysterious quale. The addition results from the illusions created by a poor grasp of the ordinary celluloid quale that existed all along. Chapter Nine takes up this issue at considerable length.

My sympathies lie on the side of silver-screen qualia. The better part of Chapter Ten will be devoted to the sketch of a metaphysical framework that is at once permissive enough to allow for the instantiation of silver-screen qualia and

rigorous enough not to give offense to the sensibilities inspired by the longing for naturalistic sobriety.

4.3.1.1 Qualia and Secondary Qualities

In undergoing the shift from the silver screen to the celluloid the history of qualia repeats the history of secondary qualities. According to one story (See Nagel (1974); and especially Searle (1992: Chapter 5)), the reduction of secondary qualities succeeded only after their appearance was distinguished from their reality. Their appearance, i.e., their phenomenal features (or qualia) were 'carved off' (Searle's term) and "swept into the philosophical dustbin of the mind." (Armstrong 1984a: 176) Thus cleansed of reduction resistant features, nothing stood in the way of identifying that which was real about secondary qualities with primary property complexes of one's choice. While not without merits this maneuver is an interim solution at best. The relocated phenomenal properties (or qualia) still await an account. And this account cannot proceed along the same lines as the previous one. The 'carve and sweep' procedure is not viable when dealing with mental properties. The 'carve' procedure fails on the grounds that *"where appearance is concerned we cannot make the appearance reality distinction because the appearance is the reality."* (Searle 1992: 122) And the 'sweep' procedure fails because it would be simply silly to attempt to solve a problem by sweeping the problematic feature into the mind where this feature is mental to begin with. (Compare Rorty (1965: 199)) More simply put, when you work in the mind you cannot do the 'sweep' because the mind has no dustbin of its own. This time around the problematic features cannot be hidden; they must be reduced. Opinions differ on whether this can be done. Opinions also differ about the consequences of irreducibility. But all parties agree that the appearance/reality distinction and the attendant 'carve and sweep' maneuver is inapplicable to qualia.

The new way of thinking about qualia ended this consensus. Both Lycan and Rosenthal apply the appearance/reality distinction to qualia. Here is how Rosenthal puts the matter:

> We pick out physical objects, and thus classify and discriminate among them, by reference to how they appear to us...Parallel remarks hold for sensory states. We classify such states by reference to what it is like to be in those states. What it is like to have a certain sensation is how that sensation appears to us. So, as with physical objects, we pick out sensory states and discriminate among them on the basis of how they appear to us. (Rosenthal 1991: 19-20)

Rosenthal simply redeploys the strategy at the mental level. The same move is apparent in Lycan who acknowledges "that the notion of an appearance/reality distinction for conscious awareness [is] odd on its face" (Lycan fc.: 25). Nev-

ertheless, he insists that the "Internal Sense theory implies an appearance/reality distinction for subjectivity." (Lycan fc.: 20) There is the way the quale is, and then there is the way the quale seems to one. And these two things are distinct, even in the case where all goes well and the representation of the quale is accurate. The upshot of all this is a manageable notion of qualia. The quale proper has been separated from its appearance thereby rendering it unmysterious.

So much, then, for a sketch of the parallel between the fate of secondary qualities and the fate of qualia. Before leaving this issue behind I want to briefly comment on it. The familiarity of the 'carve and sweep' maneuver is apt to conceal from clear view just how peculiar its present application is. This has not been entirely lost on its proponents. And they have attempted to soften the blow. But the remedy's effect tends to be even more disorienting. Exploiting the dustbin metaphor to its limits, the gist of the remedy can be put as follows. The lack of a metamental dustbin does not matter, for our mind's aren't dirty. They only seem to be. But the dirt is not real, so there is no need for a real dustbin. Less metaphorically: what is proposed is phenomenal irrealism. Our mental states don't really have the problematic phenomenal features they appear to have. These features are mere illusions. Thus the presumed problems with carving and sweeping are not genuine. Rosenthal presents a bold and lucid statement of the view:

> We are willing to deny qualitative color to physical objects because we accept that their qualitative character, however we interpret it, is merely apparent. It [the judgment that physical objects are qualitatively colored] is a verdict of commonsense intuition on which we should not rely. We can say the same for the commonsense intuition that the distinctive properties of sensory states are qualitative. The inability to relocate the qualitative character of the mental properties of sensory states gives us no reason to insist that those mental properties really do have qualitative character. We need not preserve the "element of truth" in erroneous commonsense intuitions when we become convinced that these intuitions reflect only how things appear, rather than how they really are. (Rosenthal 1991: 27)

Having brought this matter to your attention, I will let it go for now. The role of illusion in our experience of qualia will be discussed at great length in Section 9.3.

4.3.1.2 The Reaction Criterion

Many lively disputes about qualia related matters are nourished exclusively by the unrecognized ambiguity in the notion of a quale. A clear grasp of the concept of qualia that the HOR theorists employ should end these altercations. The issue concerning the reactivity criterion for visual consciousness (see page 65) is a

case in point. Consider Lycan's position on this issue. As he sees it, being visually conscious of your environment involves the registering of the relevant qualia. Appropriate reactions to environmental conditions serve him as a criterion of visual consciousness.

To the ears of the friend of traditional (i.e., silver-screen) qualia this proposal sounds preposterous. For given the traditional understanding, Lycan has just espoused the view that environmentally sensitive reactivity is a criterion for qualitative consciousness, for the fact that it is like something to be the reacting subject. And that seems false. The traditionalist about qualia will argue that the driver's reacting is neither necessary nor sufficient for her being qualitatively conscious. For one can react without qualitative consciousness. And one can be qualitatively conscious without reacting. The first claim is suggested by all sorts of empirical results concerning subjects who, to all appearances, are phenomenally unconscious in the relevant way, yet manage to exploit the 'unfelt' information to guide their behaviors. (See Kihlstrom (1987)) The second claim is suggested by contemplation of the nature of qualitative consciousness. Taken in isolation, without hookup to integrative and action guiding mechanisms it (and the subject enjoying it) just sits there passively. What this means for Armstrong's driver case is that her skillful negotiation of the road and its obstacles is not proof or her being qualitatively conscious. And to grant that she enjoys a rich phenomenology as she drives is not, by itself, enough to explain why she does not crash. (See Block (1995)). How, the traditionalist about qualia wonders, can Lycan have missed all this?

The answer is that he didn't. As he sees it, the what-it's-like dimension enters the picture only at the level of inner monitoring of the qualia-bearing mental states of the driver. The reactivity criterion is not a criterion for inner monitoring and the sort of qualitative consciousness it brings forth. All the reactivity criterion tells us is that the mental states of the driver exemplify certain properties—called qualia—that the driver exploits in controlling her vehicle. And with the substance of this claim (if not with the phrasing) the friend of traditional qualia need have no quarrel. So the disagreement is discovered to be quite devoid of philosophical significance. It all boils down to a simple difference in how the two parties use the terms 'quale' and 'qualia.'

4.3.2 Summary of the Analysis

The HOR theorists have fully and knowingly embraced the new way of speaking—for them qualia are features of the celluloid, not the silver screen.[45] Many of their readers are less innovative—for them qualia are features of the silver screen, not the celluloid. And so the seeds of confusion are sown. To unre-

formed ears the claim that you are not going to deal with qualia or phenomenal qualities and the like sounds as if you were saying that you are not going to deal with qualitative consciousness. But once one has grasped how the HOR theorists use the term 'qualia' their attempt to distance themselves from qualia related problems appears in a new light. All it amounts to is simply the claim that they are not addressing the nature-problem concerning a set of properties that are distant relatives of the properties the traditional qualophile has in mind. The refusal to address this issue does not bear on the question whether HOR is a theory of qualitative consciousness. And that HOR is just such a theory has been made very clear by Lycan and Rosenthal.

4.4 HOR and Introspectionism

At the beginning of this chapter I have made some effort to distinguish between the distinct versions of HOR. Now I am about to leave these differences behind and to focus on the shared features of the hierarchical theories of consciousness. The similarities run deeper than the differences. This suggests the integrative move. My preferred label for this core of shared doctrine is *introspectionism.* HOT and HOP theorists will deplore the amalgamation as well as the label. I will presently defend both.

4.4.1 The Label 'Introspectionism'

HOT and HOP are distinct theories of consciousness. Sometimes they are even construed as competitors in the strict sense of forming an alternative. But it is doubtful whether the two approaches exclude each other even on the most detailed level of description. (See the discussion in Section 4.4.2). We have already seen that it is not difficult to abstract away from the details and to see both versions as instances of a more general idea. Higher-level, internal representation of lower-level mental states is the engine that drives both approaches. This is what the label HOR stood for.

'Higher-level, internal representation of lower-level mental states' is a long name for something which used to be called 'introspection.' The idea that introspection lights the light of consciousness—or *introspectionism* for short—is the crucial idea upon which HOT and HOP agree. The idea is not new though it did go underground for a while during the behaviorist regime. The cognitive revolution has changed that and the appeal to inner processes is no longer frowned upon in contemporary philosophy and science of the mind. The basic idea of the hierarchically organized mind in which higher-level processes perceive or cognize lower-level mental processes is straightforward and attractive. Though the

details of the actual implementation of such a system may prove difficult the model succeeds in demystifying introspection and, consequently, introspectionism.

The HOP theorists—Armstrong and Lycan—see this connection with introspection and call the explanandum of hierarchical theories 'introspective consciousness.' But Rosenthal—our paradigmatic HOT theorist—resists this convention. Confusingly he calls this sort of consciousness 'nonintrospective consciousness' and reserves the term 'introspective consciousness' for a higher grade notion of consciousness (see p. 62). I suspect that Rosenthal's reluctance is grounded in his anxious concern to distance himself from any suggestion that internal perception might play a role in his purely cognitive theory. And the etymology of the word 'intro*spection*' certainly does suggest a perceptual process. It is much less clear, however, that the idea of introspection carries this perceptual baggage.

Baldwin's Dictionary, for example, has this to say about introspection: "Attention on the part of an individual to his own mental states and processes..." After citing Locke Baldwin continues: "Other writers speak of an 'inner sense' or 'inner perception.' The suggested analogy to 'outer sense' and 'outer perception' is misleading." (Baldwin 1901) Baldwin has spoken and the student of the HOP literature is startled. Too often one has seen Locke's name invoked to furnish the inner-sense theory with some measure of historical legitimacy—a commodity that nobody can seem to live without. It comes as a surprise that Locke's text bears out both parties of this dispute. In the brief entry on introspection, Baldwin has chosen to cite (part of) the following neutral formulation of Locke: "By *REFLECTION* then...I would be understood to mean, that notice which the Mind takes of its own Operations, and the manner of them, by reason whereof, there come to be *Ideas* of these Operations in the Understanding." (Locke 1982: ii,i,4) This way of putting it does not settle the dispute between the perceptual and the cognitive understandings of introspection. The notion of noticing is sufficiently imprecise to allow for both readings. And the matter is not helped by the fact that Locke flanks this neutral passage with two others respectively suggesting the perceptual and cognitive interpretations. When he first introduces the notion of reflection he explains it as follows:

> The other Fountain, from which Experience furnisheth the understanding with *Ideas*, is the *Perception of the Operations of our own Mind* within us...And though it be not Sense, as having nothing to do with external Objects; yet it is very like it, and might properly enough be call'd internal Sense. (Locke 1982: ii,i,4)

What more could a HOP theorist ask for? The newfound respectability that comes with discovering a great progenitor gives one pleasure. This pleasure is dampened, however, when one discovers that one's 'enemies' have an equal right to claim the great man as their forebear. Locke, it would seem, was very democratically minded when it came to furnishing HOP and HOT with historical respectability. For a few paragraphs down, when he speaks of individual differences in the extent and fidelity of introspective knowledge, Locke sounds as if he were writing an early manifesto of the HOT school:

> For, though he that contemplates the Operations of his Mind, cannot but have plain and clear *Ideas* of them; yet unless he turn his Thoughts that way, and considers them *attentively*, he will no more have clear and distinct *Ideas* of all the *Operations of his Mind*, and all that may be observed therein, than...(Locke 1982: ii,i,7)

This brief historical excursion suggests that the notion of introspection (or reflection) entered the philosophical discourse in a conveniently ambiguous state—imprecise enough to allow for an inner-sense and a purely cognitive interpretation. Thus, the talk of introspection and introspectionism no more favors the inner-sense view than it begs the question against a purely cognitive account. The use of the term 'introspection' (and its cognates) is therefore quite appropriate when addressing the common core of HOP and HOT.

That said, we should register a caveat concerning this term. I have said that Rosenthal's decision to use the term 'nonintrospective consciousness' to stand for what other hierarchical theorists call 'introspective consciousness' is apt to confuse. But his terminological decision is not unmotivated. Though talk about introspection, introspective consciousness, and introspectionism is illuminating, it can also mislead. First an obvious point: introspectionism is not a theory of introspection. It is a theory of consciousness that appeals to introspection in its explanation of consciousness. Second, it is not a theory of that which Rosenthal calls introspective consciousness—the sort of awareness of one's mental states one enjoys when one makes them the explicit focus of one's reflective attention (see p. 62). Third, the label 'introspectionism' applies to all accounts that explain consciousness in terms of introspection, even to those accounts that do not employ a hierarchical model of the introspective process.[46] That is, the class of introspectionist theories of consciousness is larger than the class of hierarchical theories of consciousness. Since I will not discuss nonhierarchical theories, I take the liberty of using the terms 'introspectionism,' 'HOR,' and 'hierarchical theories of consciousness' interchangeably. Strictly speaking, this is incorrect. But if one is careful to avoid these traps inherent in the introspection talk, there is no serious objection to casting the issue in this traditional idiom.

4.4.2 Introspectionism: A Unitary Phenomenon?

I have already stated what I take to be the underlying idea common to both HOT and HOP. The existence of this common core provides all the justification needed for an unified treatment of HOT and HOP. The decision to simply speak of introspectionism and to disregard differences may, nevertheless, seem bizarre, given the intensity of the debate dividing the introspectionist factions. Only a thorough discussion of the inner-introspectionist dispute could fully allay this worry. This I cannot do here. Instead, I will pick one central issue and show that the chasm dividing the parties is less deep than it appears.

4.4.3 Introspection: Perceptual or Purely Cognitive?

Described at a sufficiently high level of abstraction, introspectionism appears as a unified position. But viewed from a less lofty perspective, HOT and HOP present themselves as competitors and the tension between them is considerable. To those who have embraced the one or the other version of introspectionism the differences between HOT and HOP theories seem significant. The disagreements are readily visible but not superficial. The debate between the two camps is lively and, as is so often the case, to the nonpartisan observer the negative arguments on both sides seem stronger than the positive arguments on either. This leaves the critic of the hierarchical approach in an enviable position. He can look on as the warring factions of the opposition undermine each other's views, busily diminishing whatever plausibility introspectionism my initially have had. We are not quite yet dealing with an internecine feud—but one can always hope.

For the most part, the HOT camp has been the aggressor and HOP has been fighting back. I will touch upon only one thread of the much more wide-ranging discussion. It concerns the central notion of the HOP—inner perception. Rosenthal has argued that inner perception does not exist, and even if it existed, it could not serve the purpose to which the HOP theorist wants to put it. Consideration of this debate recommends itself for at least three reasons. First, it is intrinsically interesting. Second, we will see that the factions of introspectionism are not as far apart as first impressions may suggest. Third, following out this dialectic will lead us to the threshold of an issue that shall loom larger later on—the question of the location of qualia.

Introspectionism in general, and HOP in particular, presents consciousness and introspection as intimately related. Without introspection there is no qualitative feel, no phenomenology. As Armstrong has put it: "Introspective consciousness seems like a light switched on, which illuminates utter darkness. It has seemed to many that with consciousness in this sense, a wholly new thing enters the universe." (Armstrong 1978: 63) Take away introspection and you

take away all qualitative feel. Introspection is what creates and sustains whatever phenomenology we experience. Introspection is what does all the work in the production of phenomenology. This tight link between phenomenology and introspection has tended to obscure an important fact. Note that the claim here is not that introspection as such has a characteristic phenomenology. The claim is, rather, that whatever phenomenology we enjoy, we enjoy because we introspect our qualitative mental states. The pivotal role that introspection plays in the production of our visual, auditory, tactile, etc., phenomenology (according to HOP) makes it easy to overlook the quite separate question whether the inner sense has a proprietary phenomenology of its own. I suspect that this is the reason why the issues of the phenomenology of introspection and the phenomenology that introspection produces have sometimes been run into one.

HOT theorists (among others) have seen very clearly that the inner-sense story faces a phenomenological problem. Each known modality of perception comes with its proprietary phenomenology. Hence, if the HOP theorist's inner monitoring is to be understood as a form of perception, it too must sport a phenomenological dimension of its own. That is, just like vision and hearing come with a characteristic phenomenology that is uniquely theirs, we should expect our introspections to be accompanied with characteristically introspective qualia. But there is no such 'sixth phenomenology' and consequently no sixth sense. Therefore, introspection is not inner perception. HOP is not a viable option for the introspectionist.

But the friends of inner perception are not impressed by this argument. Lycan's rebuttal seems straightforward enough:

> The Inner Sense theorist does not contend...that internal monitoring is like external perception in every respect. And in particular, we should not expect internal monitoring to share the property of involving some sensory quality at its own level of operation. (Lycan fc.: 34)

Lycan has more to say on this matter. But let us pause here. If that were the whole of the HOP theorist's rebuttal we would have to agree with Rosenthal that in this case the "analogy with perception [is] idle." (Rosenthal (fc.: Section V); see also Rosenthal (1991: 31)) For the explanation: 'X is just like Y save that it lacks Y's crucial feature' may be humorous, but illuminating it is not.

Rosenthal's reply has considerable force. But it need not be the last word in this exchange. For it makes at least two assumptions, neither one of which is beyond question. First, it assumes that there cannot be perception without a corresponding phenomenology. Second, it assumes that Lycan's rebuttal depicts introspection as perception devoid of phenomenology. I will address these assumptions beginning with the second.

4.4.4 Is there an Introspective Phenomenology?

Rosenthal presents an argument that supports the second assumption, viz., that Lycan's rebuttal depicts introspection as perception devoid of phenomenology. In effect, Rosenthal poses a dilemma. He argues that there are only two plausible sources for the phenomenology of introspection. And then he shows that neither one of them can do the job. On the one hand one might hold that the qualia of introspection are identical with the sensory qualities of the introspected mental states. But Rosenthal argues that this suggestion will not do:

> Since sensory states need not be conscious, their sensory qualities are independent of their being conscious. So the characteristic quality that, on the perceptual model, being conscious introduces must be distinct from the sensory qualities that sensations already have. (Rosenthal 1991: 31)

In the paper from which I just quoted, Rosenthal is content to leave it at that. He concludes that, if the sensory quality of the sensations is not the source of the introspective phenomenology, "then it is a mystery what those new qualities could be." (Rosenthal 1991: 31) In the meantime the mystery has deepened. Now Rosenthal has upgraded his argument into a dilemma by revealing a second horn and arguing that it is no more attractive than the first. If the qualia of introspection cannot be the qualities of the lower-order state (the introspected sensation), then it appears that they have to be the qualities of the higher-order state (the introspection). But that will not do because in introspecting a sensory state "it is the quality of that state we are conscious of; so how could the sensory quality of a higher-order state explain our being conscious of the lower-order quality?" (Rosenthal fc.: Section V) This challenge is stated in so compressed a manner that no confident interpretation of Rosenthal's point appears justified. As I read it, the second horn of the dilemma is puzzling. Not that the question is a bad one—but how could somebody with Rosenthal's views see a problem here? Consider how we manage to see a tomato (according to Rosenthal). The tomato's physical redness causes us to have a sensory state that has sensory redness or a red quale—that from which the phenomenology of vision is made. It is in virtue of our being in this sensory state with its red quale that we perceive the tomato as red. Now take this story inside and consider how we manage to introspect the sensory state with its red quale. Presumably the mental state's red quale will cause us to have a introspective perceptual state that has 'introspective redness'—that, whatever it is, from which the phenomenology of introspection is made. And it is in virtue of our being in this introspective state with its introspective quale that we introspect our sensory state as red.

I fail to see the problem that Rosenthal detects in this approach. One who claims to understand the story about how we perceive tomatoes (as Rosenthal

does) has no basis upon which to ground his professed incomprehension of the story about perceiving sensory states. The analogy between the cases is too tight to allow for differential treatment. Of course there are differences. We do not, for example, know what these so-called introspective qualia are. But as a first answer we might say that they stand to sensory qualities in roughly the same relationship that sensory qualities stand to the perceptible qualities of physical objects. Rosenthal does not, of course, share this assessment. As he sees it, there is no room for introspective qualia that fit the HOP theorist's bill. And without the required phenomenology there is no inner perception. This is all he needs to close the book on HOP.

As I have tried to show, the HOP theorist has an easy way out. The second horn of the dilemma is blunt. The HOP theorist could comfortably settle down on it and hold that introspective qualia are properties of higher-order perceptual states. But Lycan will have none of that. In a perplexing move he insists on *denying* that introspection has a phenomenology. That is, he accepts Rosenthal's verdict: No proprietary phenomenology! He merely rejects the suggestion that the verdict is damning.

To advance our understanding of Lycan's position we will do well to look at the concluding part of his defense against the 'no phenomenology' objection (cited above, 80). Lycan told us that we should not expect a proprietary phenomenology of introspection. Now he tells us why this is so:

> For the sensory properties involved in first-order states are…the represented features of physical objects; e.g., the color presented in a visual perception is the represented color of a physical object. First-order states themselves do not have ecologically significant features of that sort, and so we would not expect internal representations of first-order states to have sensory qualities representing or otherwise corresponding to such features. (Lycan fc.: 14)

This passage raises more questions than it answers.[47] But on its most straightforward reading it does suggest that the second assumption—viz., that Lycan's rebuttal depicts introspection as perception devoid of phenomenology—is correct. More precisely, it suggests that introspection has no phenomenology 'on its own level.' It appears then that Lycan is determined to bite this bullet—a dangerous stunt even under the best of circumstances. But there may be a way to soften the blow. Even if introspection has no *sui generis* phenomenology, it still might come with a 'borrowed phenomenology'—viz., the phenomenology (or some derivative of the phenomenology) of the sensory states at which it is directed. William Lyons, for example, describes the situation in these terms:

> But when we come to describe our alleged introspectings, we do so in terms of what we introspect, not in terms of what it is like to introspect. Any expe-

> riential qualities in introspecting a patch of blue seem to be borrowed from the first-order experience, which is perceiving the patch of blue. (Lyons 1986: 96)

One passage in Lycan's text obscurely hints at this possibility. He writes:

> Introspection represents a first-order state under an aspect, or as being a certain way, and that "way" doubtless has *something* to do with the first-order state's own quale or sensory quality. So if Rosenthal's term "mental quality" is taken more broadly than "sensory quality" in his original sense, it is possible that...every scanning of a first-order state does "involve" some distinctive mental quality that is distinctively related to the first-order sensory quality. (Lycan fc.: 14)

Disarmingly Lycan adds that this idea needs to be 'demystified' if it is to be taken seriously. But as of yet the mystery stands. One wants to know how one phenomenology can do double duty. An attractive but naive picture suggests itself. Imagine that your perceptual states are transparent (as if they were made of glass, like a windowpane). In that case, if you introspect an (internal) perceptual state you see right through it (like through the clear windowpane), straight out of your head and onto the surface of the external object that you are seeing (with your eyes). Thus the redness of the tomato you see would be the quale of your perception and also the quale of the introspection of your perception. But, of course, it cannot really work like that.

Other writers have considered the question whether the perceptual model of introspection might be saved by the idea that introspection has a borrowed phenomenology. I'll present two passages from Fred Dretske and Sidney Shoemaker respectively. Though they start from quite different assumptions, their conclusions show a striking degree of convergence. In his penetrating discussion of introspection, Dretske (1995: Chapter 2) comments on the inner-sense theory and makes some remarks that pertain to the idea of a borrowed phenomenology. He writes:

> If there is an inner sense, some quasi-perceptual faculty that enables one to know what experiences are like by "scanning" them, this internal scanner, unlike the other senses, has a completely transparent phenomenology. It does not "present" experiences of external objects in any guise other than the way the experiences present external objects. If one is aware of experiences in the way one is aware of external objects, the experiences look, for all the world, like external objects. This is very suspicious. It suggests that there is not really another sense in operation at all. (Dretske 1995: 62)

And here is how the issue presents itself to Shoemaker:

> Now some of the states we are aware of in introspection are themselves sensory states having a phenomenal character. And in the case of these, it is natural to say that there is "something it is like" to be aware of them. But it seems plain that this "something it is like" is just the phenomenal character of the states themselves, and not the phenomenal character of still other states that are sense impressions of them...adopting the perceptual model of introspection, one thinks of "feeling" as the mode of perception by which one has introspective awareness of pain. But that has to be a mistake...There is something it is like to be in pain, or to feel pain, but there is nothing in addition it is like to be aware of pain, or of feeling pain; and the same goes for other sensory states. And so there is no such thing as a sense impression of a sensory state, having a phenomenal character of its own. (Shoemaker 1986: 114)

As even these brief passages make clear, Dretske and Shoemaker have much to disagree upon. Notwithstanding these differences they reach remarkably similar conclusions in this particular case. The idea that one might furnish introspection with a phenomenology borrowed from its object strikes them as unacceptable.

Thus we are thrown back to the hard line that Lycan endorses in the quotation with which we began. Strictly speaking, introspective perception presents mental states unaided by a phenomenological medium. In introspecting your perception of the tomato you become aware of the tomato-perception. So far so good. But now reflect on the special nature of this awareness. It is a form of perceptual awareness which is not an awareness of something as of something red, or blue, or round, or bright, or smooth, or loud...There is nothing *as which* your mental state perceptually presents itself to you. One might say (in a slightly mystical tone) that this is a case of *pure* perceptual awareness of the object. But one should say (in a resoundingly debunking tone) that this is not a case of perceiving at all.

Hard lines are often advocated in the interest of mystery reduction—away with all the woolly, touchy-feely stuff. But if pushed mercilessly, hard lines, too, can give rise to their own peculiar kind of austere mysteriousness. Lycan's hard line that introspection is perception without phenomenology may well be a case in point. On these grounds one may be led to agree with Searle when he says: "if by "introspection" we mean a special capacity, just like vision only less colorful, that we have to *spect intro,* then it seems to me there is no such capacity." (Searle 1992: 144)

Note that this line of thought has taken us back to the first assumption from above—the assumption that there cannot be perception without a corresponding phenomenology. It is now time to confront it, for it too is controversial.

4.4.5 Perception without Phenomenology?

Many authors agree that introspection has no proprietary phenomenology. But these same authors tend to also subscribe to the slogan "no perception without proprietary phenomenology." (See, e.g., Lyons (1986: 96); McGinn (1982: 51); Rosenthal (fc.: Section V); Shoemaker (1986: 114)) So they conclude that introspection is not perception. Lycan avoids this conclusion by rejecting the second premise (which is simply a statement of the first assumption introduced on page 80). What, then, is the status of the first assumption?

Many will hold that the first assumption is obviously false: there can be perception without phenomenology because there is. The phenomena of blindsight and subliminal perception show that. Unfortunately, things are less clear cut than this rebuttal would have it. Simply pointing to blindsight and other such phenomena may not be enough to settle this question. Much will turn on one's account of perception. William Alston, for example, has presented an account of perception that makes a phenomenology into a necessary condition of perceiving (see also page 119). On this theory, the borderline cases of perception—e.g., blindsight, extrasensory perception, the perception of ethical value, the perception of mathematical objects, the perception of God—must involve a proprietary phenomenology on pain of being reclassified as intuitions, hunches, mere beliefs, or whatnot. Alston is particularly interested in the mystical perception of God. To portray this as a case of genuine perception—with all the attendant epistemic credentials—Alston must reveal the phenomenological dimension of these encounters. He braves this daunting task by pointing to a nonsensory phenomenology. This idea places considerable stress on our idea of a phenomenology.[48] But an odd phenomenology is preferable to no phenomenology. For without phenomenology there is no perception. Given this sort of approach, mention of the phenomena of blindsight and subliminal perception do nothing to establish the case for the existence of perception without phenomenological dimension. Either they are nonperceptual or we have to furnish them with a proprietary phenomenology. In neither case do they support the intended conclusion.

Note, however, that this line of reasoning is no stronger than the theory of perception it presupposes. And there is little doubt that the advocates of HOP will be underwhelmed. Take for example Armstrong's view of perception. One of the most striking features of this account is what it omits. Armstrong quite deliberately refuses to make much of the specifically phenomenological dimension of perception. If we are to believe Armstrong, perception is, to put it crudely, a matter of belief acquisition. Thus, it is neither here nor there whether the blindsighted person has a phenomenology, be it sensory or nonsensory. If she forms the appropriate beliefs in the appropriate way she perceives in the full-

est sense of the term. And the same holds true of introspection. The absence of a proprietary introspective phenomenology poses no problem for the inner-sense account of introspection. To form the right sorts of beliefs about one's inner states just is to perceive these states. Given the Armstrongian view, any further complaints about phenomenology miss the point.[49]

4.4.6 Narrowing the Gap between HOT and HOP

These brief comments on the debate surrounding the first assumption do not begin to settle it. The issue quickly balloons into a full-scale discussion of perception—a task I cannot undertake here. This inconclusiveness is not limited to the first assumption. For, as we have seen earlier, the debate about the second assumption leads up to and depends upon the debate of the first assumption. For all I have said, then, the dispute about the two assumptions and, consequently, the dispute about the perceptual or nonperceptual nature of introspection is still wide open. A resolution of the conflict will turn less on specifics of the HOP and HOT as such than on a consensus on much more general questions in the theory of perception and the philosophy of mind. The discovery that the dispute is rooted in these much more general philosophical disagreements lets one see the inner-introspectionist feud in a new way. The HOT/HOP dispute may now appear more like a symptom of a deeper disagreement than as a deep philosophical debate in its own right. This result may be seen as small step in the direction of narrowing the gap between HOT and HOP.

Other authors (especially Dretske (1995); Güzeldere (1995)) have argued that the case against the perceptual interpretation of introspection is definitely stronger than I have made it out to be.[50] From this Güzeldere concludes "that the presumed distinction between the two kinds of higher-order accounts of consciousness…is unfounded as it stands. In particular, the higher-order perception accounts, taken literally, can only be a species of higher-order thought accounts." (Güzeldere 1995: 337) If Güzeldere is right the gap is not merely narrowed—it is gone. Hence the decision to speak of the unified phenomenon of introspectionism is not bizarre but demanded by the facts.

Dretske's assessment of the perceptual take on introspection is even more scathing. This is especially relevant since he then proceeds to argue that Rosenthal's HOT account, when developed sufficiently to withstand obvious objections, collapses into HOP. When faced with the problem that small children and animals are conscious but have no higher-order thoughts, HOT responds by, as it were, cheapening higher-order thoughts to the point where even conceptually challenged subjects can have them. To Dretske this move seems like cheating:

> This won't do. It is a thinly disguised conflation of higher-order thought with higher-order experience. I can see an object...without knowing what it is, but if this is the model we are to use to understand the way thought can be about experience without representing the experience *as* an experience, then our relation to our own experiences is being likened to our relation to the objects we see and hear. Experiences...must be viewed as existing in a "relevant sensory field" of their own. This, though, is an "inner sense" theory of what makes a state conscious. (Dretske 1995: 112)

Again the gap is not merely narrowed but closed. Güzeldere told us that HOP is merely HOT in disguise. Dretske tells us that HOT is merely HOP in disguise. Clearly this situation invites further clarification. But for our purposes we need not be too worried about who wears the mask. As long as either one of the competing hierarchical theories is merely a version of the other, we will be justified in lumping them together under the label 'introspectionism.'

The reflections on Armstrong's theory of perception—sketchy though they are—point towards another reason in favor of amalgamating HOT and HOP. If perception is belief and beliefs are thoughts, then perceiving one's mental states is a matter of one's thinking about them. (Compare also Güzeldere (1995: fn. 22)) If—and that is a big if—this caricature of Armstrong's theory of perception is acceptable, then the intra-introspectionist conflict is merely superficial. I conclude, then, that the decision to talk of the unitary phenomenon of introspectionism is well motivated, first impressions to the contrary notwithstanding.

4.4.7 Externalizing Qualia

One inner-introspectionist chasm has been closed. But another one, and a much more treacherous one at that, has opened up. In the attempt to come to terms with the problem created by the lack of a dedicated introspective phenomenology, the HOP camp has made a heroic move. They evicted the qualia from the head. Qualia were taken to be the properties of perceptual states that accounted for their introspective discriminability. All this was supposed to go on in the head. Now the qualia have been moved out, onto the objects of the introspectible perceptual states. These objects—tomatoes, trees, etc.—and their properties are outside the head. This is the sense in which the HOP theorists externalize qualia. Eventually I will argue that this ingenious move does not work.[51]

But to simply speak of an externalization of qualia does not do full justice to the HOP theorists's proposed revolution. As the qualia are being pushed out of the head they also undergo an unsettling change in their nature. They turn into something we always thought they were not, viz., ordinary physical qualities of external physical objects. So there are really two issues here: (i) Where are qualia located? (ii) What sorts of properties are they? More specifically: Are they just

the ordinary physical properties of external physical objects we perceive? To question (i) Lycan answers: Outside; Rosenthal answers: Inside. To question (ii) Lycan answers: Yes; Rosenthal answers: No. The disagreement is complete.

Lycan's official notion of a quale may not, right away, reveal his radical answers to these questions:

> The sense [of 'quale'] I have in mind is roughly C.I. Lewis' original sense, in which a 'quale' is the introspectible monadic qualitative property of what seems to be a phenomenal individual, such as the color of what Bertrand Russell called a visual sense-datum. For example, if S is visually healthy and looking at a ripe tomato in good light, the tomato will look red to S, and if S focusses her introspective attention on the corresponding subregion of her visual field, S will see that subregion as an individual red patch having a roundish shape. The redness of that phenomenal patch is the quale of the containing visual sensation. One registers such a quale whenever one perceives a coloured object as such. (Lycan 1995a: 249)

On reading this the heart of even the most traditional qualophile will go out to Lycan, but only to suffer a terrible disappointment when it arrives there. For when Lycan explains this passage in the paragraph that follows, all trappings of tradition crumble. We learn that "a quale is a represented property, an intentional object." It is intentional because "if it were an *actual instance* of redness present in the mind, there would have to be a mental object that really is red…which there surely is not." The sensation "represents the tomato…as having the color red." The sensation is not itself red. "Rather, the visual sensation represents a state of affairs in which an external object, or its surface, has a certain intrinsic property." And this physical redness of the tomato is the only color involved in one's perceiving it. As he puts it, "the "colors" involved in visual experiences [are] just the physical colors of represented physical objects actual or nonactual." (Lycan 1992: 17) So the qualia are outside and identical with the physical colors of the external objects we perceive.

Rosenthal denies both these claims. Qualia are inside and not identical with the physical colors of the external objects we perceive. Pondering the nature of the mental quality constitutive of the (alleged) phenomenology of the inner sense, he says:

> What might that mental quality be? In the case of our being conscious of a sensory state, it is tempting to hold that the mental quality in virtue of [which] we perceive that state is the sensory quality of the sensation in question. But this answer is theoretically unmotivated. When we see a tomato, the redness of our sensation is not literally the same property as the redness of the tomato. If being conscious of a sensory state is like perceiving that state, why should the quality involved in our being conscious of that state be the same quality as that of the state itself? (Rosenthal fc.: Section V)

Underlying this argument is a view of perception that contradicts Lycan's on both counts. First, perceiving a ripe tomato involves two sets of colors—two rednesses. The physical redness of the tomato's surface and the sensory redness of the mental state that is the perception of the tomato. Despite being both physical properties, physical and sensory redness differ in their function as well as in their physical makeup. The sensory redness of your percept has the function of signaling the presence of an external thing that is physically red. (See Rosenthal fc.: Section V) The physical redness of the tomato has no such function. And, unlike a physically red tomato, a sensorily red percept does not look red to the observing neurologist. That is why there aren't red patches in the brain of one who looks at a tomato. It is this latter sensory, intra-cranial redness—the red quale—that we discern introspectively. Thus, second, the quale is inside. Given all this as a background, it is easy for Rosenthal to make the point that the percept's quale cannot be the mysterious mental quality responsible for the phenomenology of introspection. What we need is a 'metasensory' quality. For to allow the coincidence of representing and represented quality would be to break decisively with the perceptual model.

The disagreements are deep and numerous. They concern at least the following four issues. (i) How is perception (be it inner or outer) to be analyzed? (ii) What would follow if qualia were instantiated in the brain? (iii) What sorts of qualities are qualia? (iv) What sorts of objects exemplify qualia and where are they located? No amount of papering over will make these fault lines invisible (much less will it make them go away). Though these issues run deeper than the disagreement about the perceptual or purely cognitive nature of introspection, they have received less attention.[52] So I will spend considerable time with these issues in later chapters. There I discuss the externalization of qualia—which will be referred to as the wholehearted relocation of qualia—as part of the more general thesis of the relocation of qualia. Chapter Seven spells out the arguments in favor of relocation. Chapter Eight considers the problems involved in the having of relocated qualia. In Chapter Nine I present arguments against the relocation move. It is there that I will be most directly concerned with the externalization of qualia encountered in the present section.

4.4.8 Qualia: Mental Paint vs. Physical Paint

In Section 4.3.1 I distinguished between celluloid and silver-screen qualia. In the previous section we encountered a related but different distinction. Borrowing Gilbert Harman's colorful term "mental paint" (Harman 1989), we can draw a distinction between qualia understood as mental paint or as physical paint. According to the mental-paint scenario, qualia are intrinsic properties of experi-

ences, where the notion of an experience is narrowly construed and neutral as between the physical/nonphysical distinction. The intuitive idea is that qualia are, in some sense, *internal* to the experiencing subject. On the physical-paint picture, qualia go not with the experience but with the experienced object, where this object is typically understood as an external, physical object, like, for example, a tomato. Given this construal, the red quale of one's tomato experience is identical with the red color of the tomato. So qualia are understood as external properties par excellence.

The terms in which this distinction is couched may suggest that the physical/nonphysical contrast is at the heart of this dichotomy. But this is misleading. The internal/external contrast best captures the intended thrust of this distinction. These two dichotomies are not the same, as is borne out by the fact that (i) an internal, mental-paint quale might be physical, and (ii) an external, 'physical' paint quale might not be physical (in one very good sense of the term). Representatives of position (i) are easy to find. Everyone who holds that qualia are neurophysiological properties belongs into this camp. The existence of the second possibility has been noted repeatedly and rarely defended. If you held that color qualia are sui generis and physically irreducible properties of perceived, external, physical objects, you arguably might count as a member of the second group.[53]

Put in terms of this distinction, the disagreement between Lycan and Rosenthal concerns the question whether qualia are best understood as mental or physical paint. Both agree that the paint has to be a nonMental property. The disagreement is one about where to locate these properties. Rosenthal puts them into the brain. Thus he opts for mental paint. Lycan puts them on the intentional object. Thus he thinks of qualia as physical paint.

One word about the relationship of the mental-paint/physical-paint distinction and the silver screen/celluloid distinction. The two distinctions are independent, so all four combinations are possible. That does not mean that all combinations have been chosen with equal frequency. A combination of mental paint with the silver-screen view was the most popular choice in post Cartesian philosophy. But the contemporary literature shows none of this uniformity. The silver-screen picture is rapidly loosing ground. And the physical-paint picture is gaining adherents in otherwise quite diverse camps. Lycan and Rosenthal are at one in rejecting silver-screen qualia—they both hold that only the qualia you are aware of contribute to its being like something to be you. But we have seen that this shared preference for celluloid qualia does not keep them from disagreeing on the further question of the location of qualia (or the nature of the paint).

4.5 Summary

In the first part of this chapter I defended the view that hierarchical (or HOR) theories of consciousness are theories of qualitative consciousness. The contrary impression is grounded in a profound ambiguity of the notion of qualia. Qualia may be taken as features of the silver screen or as features of the celluloid. The failure to pay attention to this distinction may lead one to misunderstand the scope of HOR. With this ambiguity out in the open it is possible to state the thesis more precisely. HOR works with celluloid qualia. Since it does not address the nature-problem, it does not tell us anything about the particular features of the celluloid qualia presupposed by the theory. But since HOR does address the having-problem, it does claim to tell us how one has to relate oneself to a (celluloid) quale in order for it to be like something to do so. This is the sense in which HOR can be seen as a theory of qualitative consciousness that addresses the having problem without addressing the nature problem.

The discussion of Armstrong's absentminded driver figured prominently in this argument. The official point of this discussion was to show that this case cannot be used to remove qualitative consciousness from the intended explananda of HOR. The more interesting point of this discussion emerged along the way: HOR may be too limited to account for the various degrees of consciousness needed to make sense of the spectrum of consciousness stretching from blindsight to full blown consciousness of the attentive observer.

In the second part of the chapter I argued that the differences between the various versions of HOR (HOT and HOP) are less significant than their similarities. The label 'introspectionism' was introduced to stand for the common core of these views, intentionally stressing the continuity of these views with earlier theories of consciousness. The argument in favor of treating as one two positions whose differences their proponents don't tire to point out is focused on the notion of an introspective phenomenology. Its absence dooms HOP, or so the HOT theorists think. The HOP theorists disagree. They argue that this absence is exactly what one should expect, given the correct account of perception. Thus the issue quickly balloons into a debate about the theory of perception. While unable to settle this bigger dispute, I am content to conclude that the debate concerning the existence and significance of an introspective phenomenology is too ill defined to establish irreconcilable differences where I see a unified phenomenon. These considerations are strengthened by the observation that other anti-hierarchical thinkers have presented their own reasons for running together what the hierarchical thinkers are so careful to separate. I conclude this chapter by mentioning a deeper divide within introspectionism. It concerns the location of qualia and will be discussed at great length later on in the book.

But the postponement of the discussion of this aspect of introspectionism still leaves a number of questions about introspectionism unanswered. In the next chapter I will attempt to lay bare the motivational roots from which introspectionism draws its intuitive strength. Why do people think that it is *obvious* that introspection must play some very crucial role in bringing about and explaining consciousness? As I will try to show, there are many questionable background beliefs that make this connection seem obvious. But on inspection these background beliefs are found to be insufficiently sturdy to prop up the alleged obviousness of the connection of introspection and consciousness. That done I turn (in Chapter Six) to a detailed critique of John Pollock's version of introspectionism.

CHAPTER FIVE

The Allure of Introspectionism

5.1 Explaining Consciousness through Introspection: Introspectionism

Much current thought about consciousness is guided by the conviction that introspection holds the answer to the problem of consciousness. Accordingly, a good number of the extant theories of consciousness incorporate, in one way or another, the central idea that consciousness can be explained by an appeal to introspection. The form of this explanation varies. Typically one tries to show that consciousness either consists in, or results from, one's introspecting one's own mental states. I introduced the label 'introspectionism' to stand for this family of views. The question I want to address now is this: What makes introspectionism so attractive? Why has it seemed that the marriage of introspection and consciousness is a match made in heaven? Why does this connection strike some as so obvious that it appears to them that something like this must be true?

My ultimate goal is to sever the connection of introspection and consciousness. I want to maintain that consciousness (i.e., qualitative consciousness) is prior to introspection and can therefore occur without it. And conversely, the existence of introspection does not guarantee the existence of qualitative phenomena.[54] But for now I will approach this goal indirectly. I will present a hypothesis about the background ideas from which introspectionism derives its attractiveness, its obviousness. And I hope that once these ideas are revealed, the attractiveness of introspectionism will dwindle. The bottom line of the hypothesis is this: the attractiveness of introspectionism is grounded on the (long since officially discarded) belief in the Cartesian transparency of the mind to itself. Introspectionism is something like an echo of a bang that now lies in the distant past.

This form of argument has its limits. Introspectionism may be true even if my diagnosis is correct and all the wells from which the obviousness of introspec-

tionism seemed to flow are dry. One can, after all, have the right idea for the wrong reasons. There is, however, no reason to believe that this is so.

But this still leaves us with a further question. If introspectionism is such an ill motivated approach why, then, did and do so many eminent philosophers accept it? To this question I will answer that introspectionism embodies an inchoate awareness of a crucial fact about consciousness, one to which some of its competitors are blind. It is the fact that qualitative consciousness consists in its being like something *for someone*. Introspectionism attempts to capture this dimension of consciousness by insisting that the facts highlighted by the 'what it's like' locution enter the world when the subject of a qualitative state appropriates it by an inner perception or a higher-order thought. On some competing views the *for-someone feature* drops out of sight entirely. I will suggest that introspectionism is right in attempting to capture the *for-someone aspect* of consciousness but goes about it in the wrong way.

5.2 Why Introspectionism Seems Obvious

We have noted that numerous theorists take the notions of consciousness and introspection to be closely associated. Why this should be so we have not yet seen. What, then, are the reasons for supposing that introspection is the right tool for advancing our understanding of consciousness? To many it has seemed that the connection between introspection and consciousness is obvious and therefore not in need of further support. It has seemed to them that the problems of consciousness and introspection are one and the same problem, or at least closely connected problems. Since I argue that the link between consciousness and introspection is largely spurious and, therefore, *a fortiori* not obvious, I have some explaining to do. I can see three considerations that seem to support the view that we are dealing with two closely related phenomena. I will argue that they are insufficient to support the view that the issues of consciousness and introspection are inseparable.

5.2.1 The Definitional Link

One powerful source for the belief that consciousness and introspection are of a piece can be traced back to the origin of the philosophical use of the term 'consciousness.' Locke, whose *Essay* did more than any other work to permanently locate the topic of consciousness on the philosophical map, established a precedent by postulating a conceptual link between the notions of consciousness and introspection. In Section 4.4.1 we have seen three versions of Locke's claim that being conscious just is a matter of one's introspecting one's own mental

states. Since one cannot introspect anything else, we can leave out the words 'one's own mental states' and simply say: to be conscious is to introspect. Accordingly, the connection between consciousness and introspection is simply a matter of definition. This Lockean *dictum* has had, and still has, a profound influence. That is not to say that there are many 'unreconstructed' Lockeans. Still, we should not underestimate Locke's influence, even if few contemporaries would go so far as to unhesitatingly endorse Locke's simple definitional link. By introducing the term 'consciousness' in the way he did, Locke started the tradition of treating consciousness and introspection as closely related topics.

While the foregoing considerations provide a sketch of a causal explanation for the apparent obviousness of the connection between consciousness and introspection, they do not, strictly speaking, provide a reason for it. Locke's definition of consciousness as introspection would be a reason for upholding the connection, if it could be shown that the definition was a good one. We should, therefore, ask ourselves whether we have good grounds for accepting Locke's definition. It may be intended to achieve one or the other of two very different goals. It might be put forward as stipulation that one is going to use the term 'consciousness' interchangeably with the term 'introspection.' We are, of course, free to stipulate the meanings of our terms as we please. But in doing this we run the double risk of isolating ourselves from the rest of the discussion and of confusing the discussion needlessly. Since there is no such thing as a false stipulative definition, the form that our disagreement with the stipulation must take is just that of rejecting the definition as undesirable, given the other uses of the term 'consciousness' that we want to account for.

On the other hand, the definition of consciousness as introspection might be intended to adequately capture the full spectrum of uses of the term 'consciousness.' If this is the spirit in which the definition is put forward, it must be regarded as a failure. For, as I will argue later on, the particular notion of consciousness that we are interested in—qualitative consciousness—is not captured by equating consciousness with introspection.

5.2.2 The Best-Explanation Connection

There are also less direct ways of bringing together consciousness and introspection. Consider, for example, the following generic version of a notion of consciousness popular among cognitive scientists: consciousness consists in the capacity to integrate and process information from various sources and to use this information in guiding one's behavior. How does this information become available for integration and processing? If one answers that introspection presents this information it will seem that an explanation of introspection is also (at

least the better part of) an explanation of consciousness. In this case, the link between consciousness and introspection is no longer definitional. Instead, introspection presents itself as the best explanation of the capacity with which consciousness has been identified.

This consideration in support of the claim of the connectedness of consciousness and introspection is solid as far as it goes. That is, if one limits oneself to some such narrowly circumscribed notion of consciousness, then this consideration does support the claim that the investigation of consciousness and introspection will be closely linked. But if one is interested in a richer notion of consciousness, one, for example, that focuses on the phenomenal dimension of experience, the preceding consideration has no tendency to support a link between this notion of consciousness and introspection. The failure to clearly separate various notions of consciousness is at the root of the view that introspection is the key to unraveling all of the knotty problems of consciousness. One correctly perceives a connection between introspection and one particular notion of consciousness and then concludes, by means of some unnoticed conceptual slippage, that understanding introspection is all that is required for understanding consciousness in its entirety.[55]

5.2.3 Consciousness and the Consciousness of Consciousness

The two preceding considerations in support of the link between consciousness and introspection (as well as my objections against them) may seem to have a superficial air about them. They are, after all, mostly concerned with the correct definition of words, a concern that many do not consider worthy of a philosopher's time. The third reason in favor of tying consciousness to introspection does not suffer from the same shortcoming. Instead, it springs from a subtle and important kind of confusion. It is rooted in a confusion of the thing known (consciousness) with the knowing of the thing (the consciousness[56] of consciousness). The two quotes that follow pinpoint the problem with admirable clarity:

> When asked the question, what is consciousness? we become conscious of consciousness. And most of us take this consciousness of consciousness to be what consciousness is. This is not true. (Jaynes 1976: 21)

> One can understand how introspection might seem to be the whole problem of consciousness because it is central to the problem of the epistemology of consciousness. Through introspection we come to know how we feel, see, and think as we do; thus it seems that if we want to know how we know we are conscious, we must find out about introspection. But introspection is none-

> theless logically distinct from the other states. This distinction is crucial to the understanding of consciousness...(Lloyd 1989: 182)

Whenever the area of investigation is our own mind the danger of committing such level confusions is particularly acute. While it is hard to confuse the tree in the quad with our cognition of it (our belief about it, or our knowledge of it),[57] matters stand differently when we want to investigate states that are, in some sense, of the same kind as the states by means of which we gain cognitive access to them. As long as the level confusion between consciousness and the consciousness of consciousness is not clearly seen, an introspectionist account of consciousness will remain irresistible.

The remainder of this chapter is devoted to an extended diagnosis and critique of the sources of this confusion. But before this campaign is launched it must be said that the appeal to confusion furnishes only a partial explanation of the allure of introspectionism. A good part of the attractiveness of introspectionism stems from the fact that it does express a deep insight about consciousness.

5.2.4 Introspectionism's Insight

In keeping with time-honored philosophical tradition I discover the views I do not share to be riddled with numerous confusions. The appeal to confusion is powerful. It also explains why such confused doctrines are accepted by intelligent people. They too are confused, of course. Like other universal explainers the appeal to confusion wears thin if overused. It is therefore reassuring to discover that the attractiveness of introspectionism is due to more than just a simple (or not so simple) confusion on the part of those who feel it. Introspectionism embodies a correct insight about consciousness but conceptualizes it incorrectly.

The ease with which consciousness and the 'consciousness' of consciousness are run together partially explains the propensity to substitute discussions of introspection for discussions of consciousness. But the inclination to change the topic does have a deeper root. It is the insight that consciousness (unlike rocks, tables, chairs, etc.) has built into it a characteristic relation toward an apprehending mind. Consciousness is always consciousness *for someone*. This is the feature I have called the 'subject involvement of consciousness' or the '*for-someone aspect* of consciousness.' The mistake consists in misinterpreting this directedness toward an apprehending mind in a manner that yokes consciousness to introspection. The temptation to enlist introspection to make sense of the subject involvement of consciousness is understandable. If one thinks of a conscious state as something wholly objective—like a marble, say—then one will be inclined to think that the subject has to do something to the state in order to make it its own. The conscious state is an inner state. So the subject's activity that

serves to make the state its own must consist in some sort of inner grasping. In other words, the subject must introspect the state in question in order to impart to it the feature that being in that state is like something for the subject in which it occurs.

5.2.5 Preserving the Insight

The questions that poses itself at this juncture is this: Can we preserve the insight and avoid the mistake? I will answer that we can. How to do that is, in a way, the main objective of the following chapters. It is therefore futile to attempt to say here, in a sentence or two, what I am going to say, in so many more words, in later chapters. But I hope to be able to convey at least a flavor of what is to come.

The phrase 'consciousness has built into it a characteristic relation toward an apprehending mind' will, undoubtedly, seem vague or even somewhat obscure. This haziness will grow when one is told that one must not accept what might be termed the 'normal' understanding of this cryptic phrase. For I will argue that the characteristic relation of consciousness toward an apprehending mind should not be understood as saying that one is always, in some sense, cognitively aware of one's conscious mental states. I will not, of course, deny that we are conscious, i.e., that we have qualia. But I will maintain that this having of qualia is not a matter or one's knowing or believing anything.

At this point I can only provide a hint of what I am trying to express by saying that consciousness has built into it a characteristic relation toward an apprehending mind. This is best done by characterizing this feature of consciousness negatively. Being conscious (i.e., having qualia) is not a matter of one's introspecting one's qualia. More generally, one's becoming conscious is not mediated by any form of presentational mental activity on the part of the conscious subject. Since we have qualia, and since nothing else presents them to us, I propose to view qualitative mental states as self-presenting states. To be conscious, e.g., to have a particular red quale, is for this quale to present itself to one. Being presented with a quale is not a matter of one's executing any mental activity, introspective or otherwise; nor is it a matter of one's believing or knowing anything. This suggests that the phrase 'consciousness has built into it a characteristic relation toward an apprehending mind' might helpfully be replaced with something along the following lines: 'consciousness makes itself felt to the conscious person.'

Such general comments about a position are typically of no use to anybody. Those unfamiliar with the position tend to not understand the generalities. Those who understand the generalities usually do so because they are already familiar

with the position. And nobody learns anything. So I will not attempt to extend these hints any further. In Section 8.7 I make a second attempt at previewing my position. At that point we will have a better understanding of some rival accounts of consciousness. This will make it easier to sketch in the rough outline of my own.

Having acknowledged the deep insight embodied in introspectionism I now return to a discussion of the level confusion underlying it. In the remainder of this chapter I will attempt to expose what seems to me to be a powerful but mistaken consideration that might incline one to accept introspectionism. In the following chapter I criticize a detailed version of an introspectionist theory of consciousness. Hopefully, the result of these investigations will then point us in the right direction in which to search for a more promising analysis of the characteristic relation of consciousness toward an apprehending mind.

5.3 The Epistemology of Consciousness and Introspection

Introspectionist theories of consciousness are rooted in a conflation of two distinct questions: (i) What is consciousness? (ii) How do we know about consciousness? If one answers the first question in terms that are adequate to the second question one arrives at an introspectionist account of consciousness. This is the insight that the quotations from Jaynes and Lloyd (see page 96) express so succinctly. In what follows I want to elaborate this claim. I will do this by considering three epistemological theses about consciousness. I argue that the route leading from these epistemologically motivated theses to an introspectionist explanation of consciousness is very short. After stating the three epistemological theses about consciousness I will lay out, as sympathetically as I can, the argument that takes you from there to introspectionism. The critique of this epistemologically motivated route to introspectionism will focus on the soundness of the argument. I will argue that both premises of the epistemic argument for introspectionism are open to serious objections. Hence, the epistemologically motivated push towards introspectionism fails.

Even if this critical attack on introspectionism should prove entirely successful, it can only be considered the first step of the project of prying apart consciousness and introspection. It is not aimed at the thesis of introspectionism itself. Instead, it is aimed at one way in which one might have hoped to support this thesis. All it can show is that this epistemically driven approach does not provide a good reason for embracing introspectionism which falls short of proving introspectionism wrong. This first step is thus quite indirect. But pointing out the limits of this argument is not to rob it of all force. The connection

between certain traditional views about epistemic access and introspectionism has deep and traditional roots. The insight that this connection provides no evidence for introspectionism should considerably weaken the intuitive appeal of the idea that introspection somehow must hold the key to consciousness. Perhaps introspection is the key. But if it is, this has to be shown. It cannot simply be taken to be obvious.

To close the case against introspectionism in a completely satisfactory manner two more steps are necessary. In a second step it must be shown that introspectionism is false, not merely unobvious. And in the third step it must be shown what to put in the place of introspectionism. Chapter Six will take us some distance towards accomplishing the second step. I try to make some more headway against introspectionism in Chapter Eight. The third step I broach in Chapter Ten. There I attempt to show how to salvage the insight of introspectionism while avoiding its mistakes.

5.3.1 Three Forms of Epistemic Access to Our Own Minds

Qualitative mental states have a distinguished epistemic career. At least since Descartes's times they have been praised for their superior knowability. Although their epistemic value has depreciated dramatically in recent times, the connection between qualitative states and special epistemic status has not vanished entirely. I think that the attractiveness of introspectionism can, to a large part, be traced back to this tradition of investing qualitative states with epistemic excellence.

The tendency to view our cognitive relationship to our qualitative mental states as a particularly intimate one has expressed itself in a number of theses, each one of which underwrites the faulty conflation of consciousness and introspection. The most salient versions of such theses are the special-knowledge thesis, the fallible-knowledge thesis, and the belief thesis. The function of these three epistemic theses is to secure some kind of extraordinary epistemic status for our cognitive relationship to our qualitative mental states. For our purposes it will suffice to provide rough versions of the three epistemic theses. The special-knowledge thesis states that we invariably possess a special kind of knowledge of our own qualitative mental states, a knowledge that is infallible, incorrigible, indubitable, irresistible, or that possesses some combination of these special features. The fallible-knowledge thesis resembles the special-knowledge thesis in holding that we invariably know about our qualitative mental states, but the fallible-knowledge thesis is weaker than the special-knowledge thesis in that it refrains from claiming some exalted epistemic status for this knowledge of ours. The belief thesis is weaker still; it gives up on the knowledge claim altogether but

holds that our qualitative mental states are invariably accompanied by a belief in their existence on the part of the subject who has them.

Minimally, then, each one of the three theses specifies (at least) a necessary condition for what it takes for one of our mental states to be qualitatively conscious. It follows from the special-knowledge thesis that no mental state of ours can be conscious unless we have a 'special' knowledge (e.g. infallible, or incorrigible knowledge) of it. The fallible-knowledge thesis has the consequence that no mental state of ours can be conscious unless we have fallible knowledge of it. And the belief thesis entails that no mental state of ours can be conscious unless we have a belief about it.

Context permitting I will lump the three epistemic thesis together under the label 'thesis of constant cognition.' I use the term 'cognition' and its cognates to refer in a generic way to the special/fallible knowledge or the mere belief postulated by the three epistemic theses. This cognition is a belief which, depending on its epistemic status, constitutes a piece of special or fallible knowledge, or an unjustified belief. I call the belief that embodies this cognition—no matter what its epistemic status may be—a 'qualitative belief.'[58]

Typically, the epistemic relation the (mature human) subject bears to her qualitative mental states is thought to be special also in the following way: the subject not only cognizes that she is in such and such a state, she cognizes that she cognizes that. That is, the sort of cognition embodied by a qualitative belief has a peculiar self-revelatory or transparent character.[59] This feature becomes important when comparing the thesis of constant cognition with the claims of introspectionism. One's failure to notice a qualitative belief postulated by the thesis of constant cognition argues its absence. Transparency guarantees that one would notice it, if it were there.[60]

Summarizing the gist of the thesis of constant cognition, we can say that our qualitative mental states are invariably accompanied by the corresponding qualitative beliefs with which we cognize our qualitative states in a fully transparent manner. Depending on the justificatory status of the qualitative beliefs, these transparent cognitions constitute special or fallible knowledge or mere beliefs about qualitative mental states.

5.3.2 The Transition to Introspectionism: The Epistemic Argument

The crucial claim to which we now must turn is this: These three otherwise quite disparate epistemological theses are united by the fact that their truth presupposes our capacity to introspect our conscious mental states. How, then, does one get from the three epistemically driven theses about consciousness to introspectionism? I see two ways in which one might try to answer this question. There is

a short answer; the connection it forges between the three theses and introspectionism is direct, powerful, and indefensible. And there is a long answer; the connection it forges is less direct, less, powerful, but more defensible.

5.3.2.1 Identity: Introspections are Qualitative Beliefs

The short answer identifies the qualitative beliefs of the thesis of constant cognition with the introspections posited by introspectionism. This identification turns the arguments against the thesis of constant cognition (which occupy the better part of the remainder of this chapter) into arguments against introspectionism. The case against the thesis of constant cognition is strong. It can be argued persuasively that there are qualitatively conscious states without accompanying qualitative beliefs (and also qualitative beliefs without corresponding qualitative states). Thus, the thesis of constant cognition is false. The identification of the introspectionist's introspections with the qualitative beliefs of the thesis of constant cognition would, therefore, yield the result that there are qualitatively conscious states without accompanying introspections (and also introspections without corresponding qualitative states). Thus, introspectionism would fall with the thesis of constant cognition.

But it does not fall: the identification of the introspections of introspectionism with the qualitative beliefs of the constant-cognition thesis is untenable. The primary role that introspectionism assigns to the act of introspecting one's qualia-bearing mental states is to make these states qualitatively conscious. Introspecting a mental state with a given quale lets you feel that state; it makes it like something to have this state. Take, for example, a toothache that begins while you are sleeping (and thus, presumably, not introspecting your ongoing mental states). Before you wake up there is in you a mental state that has a pain quale. At this point you don't feel the pain. The pain becomes felt only after you wake up and introspect it. But to say that introspecting the pain makes it like something for you to have it is not the same as to say that in introspecting the pain you elevate it into an object with a privileged and self-revealing epistemic status. As I understand the introspectionists, they would want to distance themselves from the latter epistemic claim. But the role played by the qualitative beliefs of the three epistemic theses is precisely that of providing you with this sort of special, self-revelatory cognition of your qualitative mental states. Simply put, the introspections of introspectionism produce feeling rather than cognition, whereas the qualitative beliefs of the three epistemic theses produce cognition rather than feeling. Thus, the identification of introspections with qualitative beliefs is misconceived. The concepts of these two sorts of mental states play profoundly different roles in their respective theories.

The difference in role is particularly marked as regards the feature of transparency of self-revelation. Introspections are not self-revealing. That is a crucial point in the defense of introspectionism. Otherwise it would be open to a very simple objection. As numerous philosophers have pointed out (see Section 5.9.2.1 below), we are not aware of any introspections in being conscious. To many it has seemed that this is all it takes to refute introspectionism. But, as the introspectionists patiently pointed out time and again, their introspections (unlike the qualitative beliefs of the three epistemic theses) are not self-revealing! Their invisibility to the 'inner eye' (in most cases) is exactly what introspectionism predicts. Thus, our failure to be introspectively aware of them serves to confirm rather than to refute introspectionism.

5.3.2.2 Presupposition: Qualitative Beliefs Presuppose Introspections

So much for the first answer. I now turn to the second, long and rather less exciting answer. One who rejects the identification of introspections with qualitative beliefs might see the following, more indirect, connection between the thesis of constant cognition and introspectionism. In order for one to enter into any one of these different cognitive relations to one's qualitative states, one must introspect one's qualitative mental states. For it is by introspecting them that the mind presents these states to itself, thereby making them into possible objects of belief or knowledge. Thus, one's cognizing one's qualitative mental states becomes dependent upon one's introspectively accessing these states. The special-knowledge thesis, the fallible-knowledge thesis, and the belief thesis differ only with respect to the strength they assign to this introspectively mediated mode of cognitive access to our qualitative mental states. But they are alike in making introspection into a necessary condition on our having (and so our epistemically accessing) our qualitatively conscious mental states.

The argument underlying this reasoning—the epistemic argument for introspectionism—has the following form. The first premise is provided by the thesis of constant cognition: (1) For one to have a qualitatively conscious state, it is necessary that one cognize this state. The second premise was stated (but not yet defended) above: (2) For one to cognize one's qualitatively conscious states, it is necessary that one present those states to oneself introspectively. From this it is concluded that (3) For one to have a qualitatively conscious state, it is necessary that one introspect this state. Thus, the connection between qualitative consciousness and introspection is established. It is an additional step—but one that begs to be made once this connection is revealed—to assume that introspecting a state is what makes it qualitatively conscious. This final step is of course not deductively valid but it is, nevertheless, very plausible under the circumstances. The framework of the thesis of constant cognition provides an independent rea-

son—independent, that is, of introspectionist considerations—for thinking that each qualitatively conscious state must be accompanied by an introspection. Since introspection is already on stage, it makes good theoretical sense to put it to further explanatory use—any additional explanatory work one can make it do is free of charge as it were. In this situation it is near at hand to theorize that introspection ignites the light of qualitative consciousness in the qualia-bearing states at which it is directed. This is how introspectionism about consciousness arises naturally within the setting of the thesis of constant cognition.

The resulting picture is this. Every qualitatively conscious state is accompanied by the corresponding qualitative belief by means of which the conscious state is cognized. Introspection plays a twofold role in this scheme. A mental state becomes a possible object for cognition by being introspected. And if the introspected mental state sports qualia the introspection that serves to make it into an object of cognition also makes it like something to have that mental state. So it isn't the cognition of the mental state (embodied in the qualitative belief) but the logically prior introspection that makes it be qualitatively conscious. What makes this story about qualitative consciousness so elegant and economical is that the introspective apparatus it invokes is already in place. This apparatus was introduced to account for the special epistemic status of our mental states. It so happens that this same introspective apparatus also yields an explanation of what makes these states qualitatively conscious.

This, then, is the reason for our interest in the three epistemological theses: each one of them forges a connection between consciousness and introspection. On all three theses introspection turns out to be a necessary condition of consciousness. To know infallibly, or to know fallibly, or to believe that I am in mental state *m* I have to represent *m* to myself in some manner. And, since *m* is an inner state of mine, I am able to represent *m* to myself only if *m* is the object of my introspection. In this manner each one of the three epistemic theses requires us to have introspective access to any state of ours that is to qualify as a conscious mental state.

Given these considerations, the tendency to approach consciousness via introspection does seem eminently sensible. So much for a sketch of the 'long answer' to the question how one gets from the three epistemically driven theses about consciousness to introspectionism.

5.3.2.3 Special Epistemic Status: A Straw Man?

Before I turn to a critique of the epistemically motivated support for introspectionism, I want to deflect an objection that the 'long answer' is bound to trigger. I have laid out a fairly complicated route that may lead one to accept the mistaken introspectionist interpretation of the directedness of consciousness toward an

apprehending mind. But how plausible is it to hold that the contemporary introspectionists actually followed this route? More particularly, how plausible is it to hold that they accept the special epistemic status that the thesis of constant cognition (in any of its three forms) confers on our qualitative mental states? These are rhetorical questions, for we know very well that they would accept nothing of the sort.

But I think that the background story that I have presented may still shed some light on the contemporary enthusiasm for introspectionism. In the setting of the three epistemic theses, introspectionism appears as a theoretically well motivated account. So much so, it seems, that it attains the status of obviousness. What we are now dealing with is, I believe, a phenomenon that we might call cognitive inertia. Introspectionism continues to seem obvious (it maintains it's 'obviousness momentum') even while and after the framework from which it acquired its obviousness is rejected. The difficulty involved in far reaching belief revision is partly responsible for this phenomenon. It may take years for one to identify and correct all the epistemic consequences entailed by a change of belief close to the core of one's doxastic system. A second factor that contributes to the phenomenon of cognitive inertia is best explained by means of an analogy. We are all familiar with the way in which means can gradually turn into ends. One begins to jog to lose weight. But soon the jogging becomes an end in itself. We find a similar phenomenon in the cognitive domain. A view seems obvious when anchored to a given framework. But soon the view appears obvious *simpliciter.* Once this 'internalization of obviousness' has taken place the obviousness of the view is cut loose from the framework that conferred it in the first place. This, I suggest, has happened in the case of introspectionism.

My aim was to undermine the intuitive obviousness of introspectionism by uncovering the (discarded) views in which this intuition was originally rooted. If all went well, it should now seem puzzling why one used to think it obvious that introspection must hold the key to consciousness. Note that this sort of subversive strategy does not presuppose that the currently practicing introspectionists accept the special epistemic status thesis in any of its forms. On the contrary, it tries to capitalize on the fact that these theorists abandoned the soil in which their views are rooted. That does not prove their view false. But it strongly suggests a reevaluation.

5.4 Severing Consciousness from Introspection

At this point we are addressing the question why anybody would think that introspection has something to do with qualitative consciousness. Or, put more

strongly, we are asking why anybody would think that it is obvious that introspection has everything to do with consciousness. Distinguish this question from the question whether it is intelligible exactly how introspection is supposed to bring forth consciousness. Right now we are dealing with the prior question why introspection came to seem like a promising candidate explanans in the first place. Thus the critique of the epistemological argument for introspectionism is not designed to show that introspection cannot explain consciousness. The purpose of the critique is to undermine the epistemically based reasoning that leads to the 'insight' that something along introspectionist lines must be true, however the details of the final introspectionist theory turn out in the end.

The epistemic argument for introspectionism (see p. 103) has two parts. A deductive part, (3) follows validly from (1) and (2), and a nondeductive part, the step from (3) to the introspectionist conclusion is an inductive leap of sorts. But, as far as such leaps go, it appears to be in good shape. I will therefore not question the validity (broadly construed) of the epistemic argument. Instead, I will focus on the premises of the argument in an attempt to prove it unsound. I start off by briefly discussing the introspective access assumption enshrined in premise (2). This leads naturally into a detailed discussion of premise (1). The strength of premise (1) varies according to which interpretation of the thesis of constant cognition is chosen. I will spend some time with the special-knowledge thesis and the fallible-knowledge thesis. But the bulk of the discussion will be devoted to the discussion of the belief thesis. I hope to show that the attempt to present an epistemically based support for introspectionism collapses because the three epistemic theses from which this argument proceeds are, all of them, false.

5.5 The Introspective Access Assumption

The second premise of the epistemic argument states what we may call the introspective access assumption: Introspection is the only means by which the mind can present qualitative states to itself so as to make them into possible objects of belief or knowledge. Thus introspecting one's qualitative states becomes a necessary condition on one's cognizing those states. The idea is that a qualitative states becomes a possible object of belief by being introspected. The state's being introspected brings it within the purview of our propositional attitudes. As long as the qualitative state exists unintrospected it is nothing *for* the subject. Thus, the subject is not in a position to form a qualitative belief about it.

What we seem to have here is a special case of a Russellian acquaintance principle.[61] Russell argued that "all cognitive relations—attention, sensation, memory, imagination, believing, disbelieving, etc.—presuppose acquaintance."

(Russell 1914: 127) And he explicitly included introspection as the form of acquaintance on which all knowledge about one's own mind is based. Commenting on acquaintance by introspection, he writes:

> We are not only aware of things, but we are often aware of being aware of them. When I see the sun, I am often aware of my seeing the sun; thus 'my seeing the sun' is an object with which I have acquaintance. When I desire food, I may be aware of my desire for food; thus 'my desiring food' is an object with which I am acquainted. Similarly we may be aware of our feeling pleasure or pain, and generally of the events which happen in our minds. This kind of acquaintance, which may be called self-consciousness, is the source of all our knowledge of mental things. (Russell 1982: 26-27)

A comment on Russell's terminology is in order. In contemporary introspectionist theories the term 'aware' seems to be used in an exclusively propositional sense. This usage may incline one to read Russell as saying that being acquainted with a given mental state is *ipso facto* a matter of one's having a cognitive, propositional attitude towards the state. This is not what Russell has in mind, as the following passage makes very clear:

> Acquaintance, which is what we derive from sense, does not, theoretically at least, imply even the smallest "knowledge about," i.e. it does not imply knowledge of any proposition concerning the object with which we are acquainted. ... "Knowledge about" is knowledge of propositions, which is not involved necessarily in acquaintance with the constituents of the propositions. (Russell 1980b: 151)

It appears, then, that the assumption of introspective access, as stated in premise (2) of the epistemological argument for introspectionism, is rooted in a respectable philosophical tradition.

That there should be such a tradition is, of course, no guarantee that there should not also be a competing tradition that makes it its business to deny the existence of knowledge by acquaintance. An early document that nicely illustrates the existence of both traditions is the Aristotelian Society Symposium of 1919 devoted to the question: Is there 'Knowledge by Acquaintance'? The vote is evenly split with G.E. Moore (1919) and C.D. Broad (1919) as the ayes and Dawes Hicks (1919) and Beatrice Edgell (1919) as they nayes. The dispute still goes on though it appears that the naysayers are steadily gaining ground.

Rather than attempting to solve this issue, I will ask a question that strikes me as more interesting. Let's assume, for the sake of argument, that knowledge by acquaintance exists and that knowledge by description is based on it. Given our specific concern, this translates into the assumption that our cognitions of our mental states are introspectively based (where it is understood that introspection

makes for acquaintance with our own mental states). Is it possible that premise (2) is false even if we grant all this? It would seem so. For the doctrine of acquaintance does not entail that every object about which we have (descriptive) knowledge is also an object with which we are acquainted. According to Russell, all our knowledge about physical objects is of this kind. Here is how he illustrates this point, using his knowledge of a table as an example:

> My knowledge of the table is of the kind which we shall call 'knowledge by description'. The table is 'the physical object which causes such-and-such sense-data'. This *describes* the table by means of the sense-data. In order to know anything at all about the table, we must know truths connecting it with things with which we have acquaintance: we must know that 'such-and-such sense-data are caused by a physical object'. ... We know a description, and we know that there is just one object to which the description applies, though the object itself is not directly known to us. In such a case, we say that our knowledge of the object is knowledge by description. (Russell 1982: 26)

This passage raises many interesting questions which I will not address. I quote it merely to exploit it as a model upon which to pattern a parallel case where the object in question is a qualitative state and the knowledge is the cognition of this state as postulated by the thesis of constant cognition. Perhaps I can know by description one of my qualitative states with which I am not acquainted. Two conditions need to be satisfied for this to happen. First, I have to have an unintrospected qualitative state. Advocates of the acquaintance principle should not have a problem with this possibility. Russell, for example, explicitly endorses this possibility: "We may have acquaintance with an object, without being acquainted with our own acquaintance. All objects of sense that we do not attend to seem to be instances. And it is logically evident that there must be instances, since otherwise every acquaintance would entail an infinite introspective series, which is absurd."[62] (Russell 1984: 121) Second, I have to be in the possession of truths connecting this unintrospected state with other states (with which I am acquainted) such that knowledge of these truths would allow me to infer the existence of my unintrospected qualitative state.

If these conditions are satisfied, I could cognize a qualitative state of mine without having introspected it (i.e., without being acquainted with it). Note that the description of this case did not violate the principle of acquaintance—the most plausible rationale for premise (2) of the epistemic argument for introspectionism. So it cannot simply be dismissed as question begging. If it is possible for one to cognize one's qualitative states in this manner, then one can no longer infer that a qualitative state must have been introspected because it was cognized. Our considerations seem to show that it is possible to form a qualitative belief about a qualitative state without relying on introspection to first 'serve up' this

state as a possible object of belief. One's knowledge of the state is purely descriptive, but it is none the worse for that. Thus, it appears that the thesis of constant cognition fails to guarantee that every qualitative state is introspected.

How much weight should one place on this objection to the epistemic argument for introspectionism? Moderation seems to be called for. The line of attack is clear, but the development of the argument is sketchy. I can see many good questions that one might want to ask about this attempt to enlist knowledge by description to attack the second premise of the epistemic argument for introspectionism. Two questions in particular are deserving of our attention. First, What are the truths knowledge of which lets me infer the existence of my unintrospected qualitative states? Second, Does knowledge by description qualify as a case of cognition of the sort envisaged by the thesis of constant cognition? I will not attempt to answer these (and other) questions here.[63] But as long as these questions stand unanswered the case against premise two is not closed.

Assume that all the hard questions of our opponent have been answered. Is it clear, then, that the epistemologically based case for introspectionism has lost all its persuasive force? I think that the introspectionist might want to resist the conclusion. After all, the epistemic argument for introspectionism did not pretend to be deductively valid in the first place. The last step is an inductive step. The introspectionist might maintain that the attack on premise (2) has merely shown that the first part of the argument is inductive rather than deductive. The whole argument might still be a good inductive argument for all that. Giving up the access assumption blocks the entailment (of (3) by (1) and (2)), but it does not sever all ties between the three epistemic theses and introspectionism. According to the most plausible view of the matter, the vast majority of our beliefs about our own qualitative states are based on our introspective access of these states. That is, under ordinary circumstances any one of the three aforementioned cognitive relations towards my qualitative states will be mediated by my having introspected the qualitative state in question. Thus, embracing the thesis of constant cognition strongly suggests the mistaken introspectionist account of consciousness. According to this weakened argument, introspectionism is not mandated but it does, nevertheless, appear as a plausible approach to the problem of consciousness. This is more than we want to grant the introspectionist.

We must, therefore, turn to a critique of the first premise of the epistemic argument for introspectionism. The case against premise (1) is considerably more damaging to the epistemic motivation of the introspectionist approach to consciousness. Moreover, the considerations enlisted there (see especially Section 5.7.3) will reflect back on the questions the present section raised about premise (2). Some of the counterexamples deployed against the special-knowledge thesis will be seen to closely fit the pattern described above. Consideration of these

concrete counterexamples may be the most effective way of raising doubts about the premise (2) of the epistemic argument for introspectionism.

5.6 Against the Three Epistemological Theses

The first attack on the argument that takes us from the three epistemological theses to introspectionism (see page 103) is not decisive. Our opponent may want to hold on to the problematic access assumption. And even if he is willing to give up the strong case for introspective access, he can try to defend a weakened version of the epistemic argument for introspectionism. In order to avoid unnecessary controversy in what follows, I will not assume that the attack on premise (2) of the epistemic argument was successful. To further strengthen the case against introspectionism, we will, therefore, have to turn to a scrutiny of the first premise. Premise (1) expresses the thesis of constant cognition—every conscious mental state is cognized in one form or another. If the three epistemological theses are wrong, the epistemically based support for introspectionism collapses, no matter how the discussion concerning the access assumption turns out.

5.7 The Special-Knowledge Thesis

First a word about how this discussion proceeds. I anticipate the following objection: If the belief thesis is false, the special-knowledge thesis is false too; so why waste time with the special-knowledge thesis in the first place, if the belief thesis is going to be attacked anyway? There is some justice to this protestation. In reply, I would say this: it is interesting to see the special-knowledge thesis fail on its own demerits, so to speak. It is worthwhile to focus on those failings of the special-knowledge thesis that are characteristic of it, i.e., those it does not share with the belief thesis. These particular failings have, after all, generated an enormous amount of philosophical heat. As Hector-Neri Castañeda has pointed out, flogging a dead horse, especially a widely influential dead horse, need not be a superfluous thing. For, first the horse might not be quite dead "and even if it were totally dead, a careful autopsy of it is the best means I know of to learn about the anatomy of…the indigestible data that would have caused its death." (Castañeda 1979: 2)

The tradition of making extravagant claims concerning the epistemic status of our beliefs about our conscious mental states is long and distinguished. The epithets attached to the special brand of knowledge allegedly involved are nu-

merous. Such knowledge has been called incorrigible, indubitable, infallible, irresistible. What is common to all these brands of purportedly special knowledge is the claim that something or other is impossible: it is impossible for us to doubt such knowledge claims, impossible for us to be corrected by someone else concerning such knowledge claims, impossible that we be wrong about such knowledge claims, impossible for us not to possess this knowledge in the relevant situations. The term 'special-knowledge thesis' refers, in a generic manner, to all of these kinds of privileged epistemic status with which our beliefs about our conscious mental states have been credited. Arguments against one or more versions of the special-knowledge thesis have been put forward by D.M. Armstrong, Keith Lehrer, and John Pollock, among others.

5.7.1 The Argument From Materialism

Armstrong's motivation for attacking the special-knowledge thesis lies in his commitment to a materialist theory of the mind:

> I wish to defend the thesis…that mental states are, as a matter of fact, states of the brain. Now if I accept the existence of introspection, as I also do, then I must conceive of both introspection and the objects of introspection as states of the brain. Introspection must be a self-scanning process in the brain. That it is logically possible that such a self-scanning process will yield wrong results is at once clear, nor is it possible to see how such a self-scanning process could yield a *logically* privileged access. So if introspection is incorrigible, or if we have logically privileged access to our own mental states, it seems that a materialist doctrine of mental states must be false. (Armstrong 1963: 418-419)

Armstrong presents this merely as his motive for challenging the special-knowledge thesis. But what he says here may just as well be construed as an argument from materialism to the falsity of the special-knowledge thesis: if materialism is true, then our mental states cannot have a privileged epistemic status. Materialism is true. Therefore, our own mental states do not enjoy a privileged epistemic status. At the time of Armstrong's writing (1963) this argument would, perhaps, more likely have been seen as a *reductio ad absurdum* of materialism than as a sound argument against the special-knowledge thesis. But the climate has changed and today this argument is widely perceived as a formidable challenge to the claim that our conscious mental states have privileged epistemic status.

5.7.2 The Argument from Conceptualization

Armstrong and Lehrer both propose arguments against the special-knowledge thesis that start from the following observation: knowledge, belief, and, in general, apprehension of anything in any form, involve conceptualization. I.e., we cannot apprehend anything without applying concepts, without classifying objects as falling under certain concepts. Armstrong takes this to show that the special-knowledge thesis (at least in one of its forms) has to be given up:

> Now, surely, the notions of classifying and misclassifying are co-ordinate notions; surely the one can apply only when it is meaningful to apply to other? We can apply a certain concept to our experience only if it is possible to withhold that concept. Yet, according to the doctrine of incorrigibility, the application of any concept except the concept we do apply is logically impossible. (Armstrong 1963: 422)

Lehrer relies on this same consideration in order to show that it is not even possible to construe noncomparative appearance beliefs ('I seem to be appeared redly to') as self-justifying, fallible, basic beliefs. *A fortiori* the argument will be directed against the special-knowledge thesis. What the above consideration shows is that a person's being appeared to in a particular way is not sufficient for her to be completely justified in believing that she is appeared to in this way. She needs some independent information about the conditions for the correct application of the concepts involved. That is, she needs the information that this is one of those situations in which her visual appearance is properly classified as of the red kind. On this route one can also arrive at the conclusion that at least a large class of our conscious mental states do not have the privileged epistemic status claimed by the special-knowledge thesis.[64]

5.7.3 The Argument from Deviant Reasons

Another argument against the special-knowledge thesis is based on the following consideration. Take a belief about one of your own present conscious states, more particularly, one of your appearance states. What epistemic status this belief has will depend on your reasons for holding it. The problem we encounter is this: the very same beliefs that would qualify for some kind of special epistemic status, when held on introspective grounds, might lose this status, when held on other grounds. Now, both Lehrer and Pollock claim that it is always possible to hold such beliefs on the wrong (i.e. non-introspective) grounds. Therefore, these beliefs lose their claim to special epistemic status. Here is one of Lehrer's examples:

> Whatever one can believe as a result of introspection, one can instead believe as a result of inference; and the inference can be based on false premises. If a woman believes she is in pain as a result of feeling pain, then, of course, she will be correct in believing that she is in pain. If, however, a man believes that he is in pain because some scientific or religious authority figure tells him that he is in pain, then what he believes may well be false. (Lehrer 1990: 54)

Pollock points out that you may end up holding an appearance belief, not on introspective grounds, but on an inductively based hypothesis, in which case your belief would lose its claim to special epistemic status. Using as his example the belief that you are being appeared to redly, Pollock puts the point as follows:

> The correct candidate for incorrigibility is a belief you can only have by being "directly aware" of being appeared to redly. The problem with the belief that you are appeared to redly is that you can hold it for "indirect" inductive reasons. (Pollock 1986: 61)

The discovery that we might hold our appearance beliefs for the wrong reasons undermines their claim to privileged epistemic status. Once one discovers that there are counterfeit appearance beliefs around, the guarantee of highest epistemic quality expires.

5.7.4 Assessment of the Special-Knowledge Thesis

Each one of these arguments has generated much debate. Nevertheless, one can witness something like an emerging 'modern orthodoxy' according to which the special-knowledge thesis has to be abandoned in the light of the reported arguments.[65] Thus, these arguments show that the claim that we cognize all our qualitative mental states cannot be construed as the thesis that our conscious mental states enjoy some kind of special epistemic status. Infallibility, incorrigibility, indubitability, and irresistibility are too strong interpretations of the thesis of constant cognition.

This result lays to rest one attempted way of linking consciousness to introspection. The friend of introspection may have wanted to argue as follows: We do know our conscious mental states in a special way. We cannot know our mental states in this special way without introspecting them. Therefore, consciousness is intimately related to, and dependent on, introspection. In reply to this argument we can now maintain that we do not have the special knowledge of our conscious states that the argument assumes. Therefore, the argument does not offer sound support for its conclusion that introspection is necessary for consciousness.

Of course the friend of introspection can easily amend his argument so as to sidestep our objection. All we have shown, so far, is that we do not have a spe-

cial kind of knowledge of our own conscious mental states. But the appeal to special knowledge is not an essential feature of the argument that links consciousness to introspection. The argument could, without loss of cogency, be restated in terms of fallible knowledge. It is to the consideration of this modified argument that we now have to turn.

5.8 The Fallible-Knowledge Thesis

The thesis that our conscious states enjoy a special epistemic status had to be given up. Is the weaker thesis that we invariably know about our conscious states defensible? I will now argue that the fallible-knowledge thesis is indefensible as well. This time, I will base the argument on the claim that the even weaker belief thesis is false: it is false that our conscious mental states are invariably accompanied by a belief in their existence on the part of the subject who has them. And since believing is a necessary condition of knowing, the falsity of the belief thesis entails the falsity of the knowledge thesis.

Unlike in the case of the special-knowledge thesis, we lose nothing of special interest by lumping together the fallible-knowledge thesis with the belief thesis. Let me back up this claim with some remarks about the relationship of these two theses. The fallible-knowledge thesis and the belief thesis are much closer relatives than may be apparent on first sight. Note that, given a certain kind of reliabilist account of justification, the fallible-knowledge thesis and the belief thesis will simply collapse into each other. If my belief that I am in a given qualitative state is reliably coupled to my having of the qualitative state in question, I am *ipso facto* justified in holding my belief. Barring certain undermining conditions, I can therefore be said to know that I am in the relevant qualitative state. But it does not take an externalist approach to justification to highlight the proximity of the belief thesis and the fallible-knowledge thesis. Internalism has a long tradition of insisting on our epistemic right to our beliefs about our own qualitative mental states. If the belief thesis is true and we do invariably form beliefs about our occurrent qualitative mental states, and if the internalist contention that we are within our epistemic rights in accepting such beliefs is correct, then it turns out that these beliefs will be true and justified beliefs. Barring certain undermining conditions, I can therefore be said to know that I am in the relevant qualitative state.

Both externalist and internalist considerations suggest a convergence between the fallible-knowledge thesis and the belief thesis. Since we are interested in these two theses only insofar as they serve as a vehicle for the advancement of introspectionism, I will not pursue the question of their epistemological relation-

ship any further. For our purposes we may rest assured that we will miss nothing of importance by skipping the fallible-knowledge thesis and turning directly to the belief thesis.

5.9 The Belief Thesis

Can the claim that consciousness has built into it a directedness upon an apprehending mind be cashed out as the belief thesis—the thesis that conscious mental states are partly constituted by (or necessarily accompanied by) a belief about them on the part of the subject who has them?

While much weaker than the special-knowledge thesis and weaker than the fallible-knowledge thesis, the belief thesis still can serve the purpose of establishing a connection between consciousness and introspection. One cannot believe that one is in a particular mental state without representing this state to oneself. And in order to represent this state to oneself one must be able to introspect this state. In this way, the weakest of the three proposed interpretations of the directedness of consciousness upon an apprehending mind succeeds in linking the issues of consciousness and introspection.

Before proceeding to criticize the belief thesis, we need to get a clearer sense of what the thesis asserts. I already indicated that the belief thesis specifies a necessary condition on one's having a conscious mental state: no mental state of ours can be conscious unless we have the relevant qualitative belief about it. But one might want to hold that the belief thesis makes a stronger claim than that. Having the relevant qualitative belief might be both a necessary and a sufficient condition upon our having the corresponding conscious mental state. A number of philosophers have held that the connection between conscious mental states and qualitative beliefs is tighter than that expressed by the conditional or the biconditional. Conscious mental states and the corresponding qualitative beliefs were either seen as identical (simple identity thesis) or as standing in the part/whole relationship (containment thesis). On the simple identity thesis the conscious state is identical with your belief that you are in this state. The containment thesis sees the conscious state as a constituent of the belief that you are in this state.[66] And, finally, one might want to understand the belief thesis as the functional identity thesis advanced by Sidney Shoemaker (see, e.g., Shoemaker (1975: 189)). It amounts to the claim that the functional identity of a conscious state type is partially constituted by the fact that it typically gives rise to a belief that one presently is in a conscious state of this particular type.[67]

On which of the various theses about the relation between our conscious mental states and our qualitative beliefs about them should we focus our critique?

We are interested in refuting those theses inasmuch as they present a means of tying consciousness to introspection. Since even the weakest form of the belief thesis manages to establish this connection, it will have to be the target of our investigation. The major task ahead of us is, therefore, to show that tokening the relevant qualitative belief is not a necessary condition on our having the corresponding conscious state. We can do that by showing that we can, and do, have qualitatively conscious mental states that are unaccompanied by correlative qualitative beliefs. I will put this point by saying that we have detached qualia—detached, that is, from the qualitative belief that one presently is in a qualitative state of this kind. Our main objective is, therefore, to defend the thesis of detached qualia. But before taking up the case for detached qualia, I want to briefly investigate the thesis of detached qualitative beliefs. I will start with the thesis of detached qualitative beliefs because this is the easiest way to prove false the most natural construal of the belief thesis. I suppose the belief thesis is most naturally construed as a biconditional: having the relevant qualitative belief is a necessary and sufficient condition upon our having the corresponding conscious mental state. If there are detached qualitative beliefs, then the sufficiency claim is false and the proponent of the belief thesis will have to adopt the more modest interpretation according to which it merely states a necessary condition on our having a conscious mental state.

5.9.1 Detached Qualitative Beliefs

The work that needs to be done to establish that there can be detached qualitative beliefs has already been done in the discussion of the special-knowledge thesis. If the arguments against the special-knowledge thesis are correct, then it is possible for us to believe falsely that we currently are in a particular qualitative state. The argument from deviant reasons (see Section 5.7.3) can serve to illustrate this claim. Because it is possible that you come to hold a qualitative belief in an indirect manner, i.e., without being directly aware of the qualitative state, it is possible that you come to hold such a qualitative belief without it being the case that you currently are in such a state. Both Pollock and Lehrer have provided plausible examples of such occurrences.

The arguments against the special-knowledge thesis may be felt to have a precious air about them. Consequently, one may experience some discomfort in making them the basis for one's acceptance of the thesis of detached qualitative beliefs. But it is easy to reassure oneself of the robustness of the intuition that qualitative beliefs might occur unaccompanied by the corresponding qualitative states. A number of old and new problems in the philosophy of mind trade on our willingness to countenance the possibility that we are now considering.[68]

Were we not inclined to take the thesis of detached qualitative beliefs seriously, neither the problem of absent qualia nor the problem of other minds would have any grip on us. Consider the problem of other minds, for example. One particularly insidious version of this problem arises, when one realizes that it is possible to doubt that one's conspecifics are conscious, even after one has satisfied oneself that their behavior resembles one's own in all relevant respects. The doubt does not concern their information processing capacities, their intelligence, their beliefs, or their qualitative beliefs. For obviously they do possess such capacities, otherwise they could not have engaged in the sophisticated behaviors we observed. And obviously they do express qualitative beliefs and proceed to act on them. What worries us is that they might be unconscious automata. We wonder whether their inner light is turned on or whether there is darkness inside. (See Dennett (1987a: 161)) In short, we wonder if they have got qualitative mental states; we wonder whether being them is like something for them. Thus, I submit, that if the problem of other minds has any grip on you, then you are already assuming the thesis of detached qualitative beliefs describes a genuine possibility. Similar remarks apply to the problem of absent qualia.

Accepting the thesis of detached qualitative beliefs amounts to giving up the sufficiency claim: having a particular qualitative belief is not a sufficient condition for one's being in the corresponding qualitative state. Therefore, the proponent of the belief thesis will have to give up the claim that this thesis is best interpreted as specifying a necessary and sufficient condition on consciousness. But he can still defend the necessity claim; and this is all that is needed to forge the connection between consciousness and introspection. For, according to this interpretation of the belief thesis, one cannot be in a conscious mental state without believing that one is in this state. One cannot believe that one is in a particular mental state without presenting this state to oneself. To present one of one's own mental states to oneself one must introspect it. And one cannot introspect it without thereby making that mental state conscious. This is a rough sketch of the short route that leads from the claim that no mental state of ours can be conscious, unless we have a qualitative belief about it, to the conclusions that consciousness and introspection are tied together inseparably. We must therefore turn to a defense of the thesis of detached qualia.

5.9.2 Detached Qualia

In this section I will consider two reasons for the view that conscious mental states can and do occur without a corresponding qualitative belief. First, I will maintain that a phenomenological examination of the nature of our experience shows the existence of qualitative mental states, unaccompanied by a corre-

sponding qualitative belief. Second, I will argue that the considerations of the nature of animal experience affords us a reason to accept the thesis of detached qualia and hence to reject the belief thesis.[69]

5.9.2.1 The First Appeal to Phenomenology

The first thing I would reply to the proponent of the belief thesis is that my experience shows the thesis to be wrong. A simple appeal to phenomenology—to the nature of my inner life—should suffice to silence the opponent. Being conscious is not a matter of having various qualitative beliefs; it is a matter of having (or enjoying, or experiencing, or going through, or feeling) various qualitative mental states. It is a matter of having qualia, not a matter of believing that one has qualia. We do, undoubtedly, have many qualitative beliefs. But the vast majority of our conscious mental states are not acknowledged in this manner. They pass through the mind unnoticed by our cognitive faculties. We do not believe, much less do we know, that we currently have these qualitative states. They manifest themselves simply as feelings, as experiences, as raw feels; no metamental activity is involved. That this is so is amply evidenced by a moment's reflection on what goes on in one's own mind. Or so one might have thought.[70]

Taken by itself, the appeal to phenomenology is not properly viewed as an argument against the belief thesis. What the appeal to phenomenology can do, at best, is to support a premise of an argument against the belief thesis. The argument is simple enough: if the belief thesis is true, then there are no detached qualia. But there are detached qualia, as is evidenced by the appeal to phenomenology. Therefore the belief thesis is false. The first premise is uncontroversial. Hence the opponent of the belief thesis will have to concentrate on making the second premise plausible. Since I cannot see an argument to the effect that there can be detached qualia, I assume the support for the second premise has to come from observation, albeit inner observation.

The significance of scientific observations garnered in an experimental setting depends, to a large extent, on the interpersonal repeatability of these observations. Something similar is true of phenomenological findings. An isolated phenomenological report can be discounted with impunity. But to disregard or contradict a widespread phenomenological report is no more reasonable than to suppress available but uncomfortable experimental results. We can, therefore, strengthen the appeal to phenomenology by documenting that it is, indeed, widespread. To do this properly would be a major project. I cannot begin to undertake this project here. But I can at least gesture towards a small number of philosophers whose phenomenological findings are in line with the one reported above.

The authors I will cite in my attempt to buttress the phenomenological finding reported above come to this issue from varying angles and report their findings in terms that differ from those I have chosen. Nevertheless, it will become obvious that there is a unifying theme running through their reports. All of them, or so I will maintain, draw the distinction between qualitative states one the one hand and qualitative beliefs on the other. And all of them claim that we can, and sometimes do, experience qualitative states that are not accompanied by the corresponding qualitative beliefs. Thus, I will enlist them all to support the thesis that I have dubbed the thesis of detached qualia.

Witness 1. Fred Dretske comes to the issue via a discussion of perception. For years Dretske has been defending his notion of simple seeing. There is, he maintains, a way of seeing that is "logically independent of whatever beliefs we may possess." (Dretske 1969: 17) He goes on to elaborate this claim in the following manner:

> What I wish to reject is the move from (1) *S* sees *D* as *Y* to (*a*) *S* believes himself aware of something *Y*-like, (*b*) *S* believes something looks *Y*, or (*c*) *S* believes something is appearing as *Y*. I am quite willing to admit that, in a certain sense, *D* must look some way to *S* (*not* to be read as: must look *like* something to *S*) in order for *S* to see *D*, but I do not think it is a necessary consequence of this admission that *S* must thereby believe something of the form (*a*), (*b*), or (*c*), or any modification of these. For *S*'s seeing *D* as something, in so far as it follows from *S*'s seeing *D*, must *itself* be understood as lacking positive belief content. (Dretske 1969: 9)

Dretske has encapsulated his view in a memorable slogan: "Total ignorance is not a sufficient condition for total blindness." (Dretske 1969: 17)

Witness 2. William Alston's overall account of perception is strikingly different from Dretske's. Nevertheless, he too is led to endorse the thesis of detached qualia in the course of defending his theory of perception—the Theory of Appearing. Alston insists that there is such a thing as that which "from the side of the subject is called *direct awareness* and from the side of the object is called *presentation, givenness,* or *appearance.*" (Alston 1991: 37) And of this cognitive relationship he says that it "is essentially independent of any conceptualization, belief, judgment, or any other application of general concepts to the object, though it typically exists in close connection with the latter." (Alston 1991: 37) Accordingly, he holds that one's being appeared to in a certain way has nothing to do with one's believing something about the way in which one is being appeared to. He says: "For something to look red to me is not the same as for me to *take* it to be red, *construe* or *conceptualize* it as red, even though its looking red typically evokes those reactions if the subject has the conceptual equipment to so react." (Alston 1991: 37) As Alston sees it, this nonconceptualized, non-

cognitive, nonjudgmental form of appearing is characteristic of sensory experience and "it is this feature of perception that clearly distinguishes it from just thinking about an object, remembering it, or fantasizing about it." (Alston 1991: 38)

I take this to show that Alston construes the paradigm case of perceptual consciousness—viz., something's looking, smelling…so-and-so to S—as a case of S's enjoying detached qualia. For, under such circumstances, something appears to one to have a particular sensory quale without it being the case that one entertains an accompanying belief about the quale or about one's experience of the quale. And, as Alston makes clear, this is not a freak phenomenon on the outskirts of the perceptual domain; no, it is *the* feature that transforms a conscious episode into perceptual consciousness.

Witness 3. John Pollock comes to the issue as an epistemologist. His critique of foundationalism (as a theory of justification for empirical beliefs) is based on his insight that the majority of our appearances (qualitative states) are *not* accompanied by corresponding appearance beliefs (qualitative beliefs). Therefore, the foundationalist is in the embarrassing position of having too few basic beliefs to support a reasonably large number of nonbasic beliefs. Were the belief thesis correct, the foundationalist would not face the problem with which Pollock confronts him. Here is how Pollock puts his radical claim: "We rarely have any beliefs at all about how things appear to us. In perception, the beliefs we form are almost invariably beliefs about the objective properties of physical objects—not about how things appear to us." (Pollock 1986: 61) He elaborates this claim by arguing that it is not only the case that we are rarely aware of such appearance beliefs, but that we have reason to think that we usually do not have such beliefs:

> If we were only unaware of having such thoughts because we do not attend to them, then all we would have to do to become aware of them is to direct our attention to them. But that seems wrong. It is not that easy to become aware of how things appear to you. It *is* fairly easy to turn your attention inwards and focus it on your sensory experience, but that does not automatically either generate or reveal beliefs about that sensory experience. To explain what I mean, consider the distinction between sensing something and forming beliefs about it. Suppose, for example, that I happen upon an abstract painting that I find very attractive. I may spend some time gazing at it, "drinking it in" perceptually. I am sensing the patterns in the painting, but if they are quite abstract and irregular so that they are not easily categorized, I may not be forming any beliefs about those patterns. This illustrates that consciously sensing something does not *automatically* involve forming beliefs about it (although I should think that in *most* cases of consciously sensing something we do form beliefs about it). (Pollock 1986: 62-63)

I do not want to go along with everything that is stated in, or implied by, Pollock's report of what he learned from the phenomenological examination of his inner life. But, as far as our central question is concerned, he makes the relevant points with great force and persuasiveness. There can be no doubt about the fact that Pollock endorses the view that an appeal to phenomenology can and does show that not all appearances are accompanied by corresponding appearance beliefs, or to put it in our language, that not all qualitative mental states are accompanied by the relevant qualitative beliefs. In Pollock we have found another philosopher who accepts the thesis of detached qualia on the basis of a phenomenological insight.

Witness 4. Alvin Plantinga's reflections on the epistemology of perception present a view very similar to the one we just saw at work in Pollock's thought. Plantinga writes: "The connections between experience and belief are multiply contingent. First, one can have the sensuous experience without forming the perceptual belief at all: one can simply look at the scene and enjoy the play of sensuous experience, forming no perceptual belief at all…" (Plantinga 1993: 92) And he follows this up with a claim about experience and experiential belief that is even more pertinent to our discussion:

> Indeed, in the typical case [of experiencing] I do not form a belief about my experience at all. It is wholly obvious, I suppose, that under these conditions we do not form beliefs describing the experience in question in 'qualitative' terms—thinking, for example, such a thing as that I am now being appeared to by way of a highly reticulated pattern of such and such a nature, color, hue, or shape: most of us who lack artistic training are quite incapable of forming such beliefs, or at least of forming beliefs of this sort with any sort of accuracy or completeness. (Plantinga 1993: 93)

Expressed in my terms, this amounts to an emphatic assertion of the thesis of detached qualia.

Witness 5. In his book *Thought and Experience* Peter Hess has made the crucial distinction between "the cognitive and noncognitive aspects of the mental" (Hess 1988: ix) into the centerpiece of a wide ranging discussion of questions in epistemology and the philosophy of mind. That is, Hess does not invoke the distinction somewhere along the way in order to solve a philosophical puzzle; rather, he begins with this distinction and attempts to develop a coherent philosophical position that is grounded in this distinction. He, more than any other writer, has heightened my sensitivity to the distinction between qualitative states and qualitative beliefs. A number of positions that I will defend later on have been inspired by Hess's careful use of the distinction and by his insistence that the occurrence of the noncognitive aspect of the mental is not tied to the occurrence of the cognitive aspect. Hess's distinction between the cognitive and non-

cognitive aspects of the mental roughly coincides with the distinction between sensations or experiences on the one hand and propositional attitudes on the other. Thus, he can also put his point by saying that the all-important distinction is that "between the sensations (John is in pain) and cognitive states (John believes that he is in pain)." (Hess 1988: 34) The following passage elaborates the distinction a little more fully and also introduces the term 'self-presentation' that I will make much of in later chapters:

> It is one thing to have an experience (to apprehend a self-presenting state), and another to be cognizant of the fact (or to believe) that one has such an experience. These two things often go together, but that does not do away with the fact that they are distinguishable one from the other. The state of affairs which consists of somebody's being in pain is not the same as that which consists of his belief that he is in pain. We cannot 'overlook' our pains in the sense that we cannot fail to apprehend them, but we certainly can 'overlook' them in the sense that we can simply fail to entertain any belief about the experience which we then have. Thus it is certainly not true that to have an experience is to be cognizant of the fact that one has such an experience. (Hess 1988: 126)

We have found another philosopher who defends the thesis of detached qualia on the basis of phenomenological insight. Witnesses could be multiplied, but I will resist the temptation.[71] It would seem, then, that the existence of detached qualia is a robust and important phenomenological fact.

In the present chapter I have appealed to this fact in an attempt to undermine what I take to be the most seductive motive for introspectionism.[72] As I have presented the matter, introspectionism is not grounded in a phenomenological finding.[73] Introspectionism is best viewed as a thesis that derives such plausibility as it has from certain claims about our epistemic access to our own qualitative states. If the claim that we have a peculiarly intimate cognitive relationship to our own qualitative mental states is understood as the special-knowledge thesis, the fallible-knowledge thesis, or the belief thesis, it follows that every qualitative state of yours must be accompanied by a corresponding qualitative belief. And once this result has been established, introspectionism is only a few short steps away.

More than one route leads from the belief thesis to introspectionism. The introspectionist may want to argue that the fact that every qualitative state of ours is accompanied by a corresponding qualitative belief constitutes the fact that we introspect all our qualitative states. Alternatively, he may hold that the fact that every qualitative state of ours is accompanied by a corresponding qualitative belief presupposes that we introspect every qualitative state of ours. Both of these arguments deliver what the introspectionist needs: a tight connection between

qualitative states on the one hand and introspection on the other hand. Once this point has been reached, there is no longer any obstacle for the introspectionist's explanation of qualitative consciousness in terms of introspection. For now the introspectionist can argue that qualitative consciousness just consists in, or results from, the fact that we introspect all our qualitative states.

By attacking the belief thesis on phenomenological grounds, I have attempted to block this route to introspectionism. What the phenomenologically established fact of detached qualia shows is that even the least ambitious of the three epistemic theses—the belief thesis—on its least ambitious interpretation—interpreted as specifying a necessary condition—is untenable. Thus, the most promising route to introspectionism is blocked.

There may, however, be other routes to introspectionism. Therefore I cannot claim to have refuted the introspectionist thesis about qualitative consciousness. Nevertheless, I hope to have put the introspectionist in a very uncomfortable position. I have robbed him of what I take to be the most promising support for his position. It is now incumbent upon him to show on what other grounds his thesis recommends itself to us.

5.9.2.2 Simple Minds

Up to this point the argument against the thesis of introspectionism proceeded in an indirect manner. I argued that the belief thesis (or one of the stronger variants of this thesis) makes the connection of consciousness with introspection seem plausible. But since the belief thesis (and *a fortiori* its stronger relatives) does not stand up to scrutiny, the basic idea of introspectionism is robbed of its support, and introspectionism does not merit further consideration.

In the argument from simple minds we have another argument against introspectionism. The argument can be formulated in a strong and in a weak version. Given the strong reading, the argument from simple minds is a direct attack on introspectionism. According to the weaker reading, it represents a further instance of the more roundabout strategy that was pursued so far. Both versions of the argument from simple minds are interesting. I suspect that the stronger version (which attacks introspectionism directly) is harder to defend than the weaker version (which attacks introspectionism via a critique of the belief thesis).

For simplicity of exposition I will first present a generic version of the argument. This will serve to get the general idea across. Then I will indicate how this generic version can be interpreted as the strong or the weak version of the argument respectively. Finally, I will express my reasons for not wanting to put too much weight on either version of the argument from simple minds.

If introspectionism is intended as an unrestricted account of consciousness in general (i.e., not merely as an account of consciousness in human beings) then it

must be true of *all* conscious beings, animals included. Mindful of "The Messes Animals Make in Metaphysics,"[74] we may be led to expect that animal consciousness will cause trouble for hierarchical theories of consciousness. Simply put, theories of consciousness that invoke metamental activity or higher-level processing seem too sophisticated as an account of consciousness in animals.

The argument runs as follows: Animals enjoy qualitative consciousness. But most (or all) animals are incapable of the relevant forms of higher-level cognitive processing. Therefore, (animal) consciousness without corresponding higher-level processing exists. Thus, the link between consciousness and higher-level processing is broken and we have one more reason to reject introspectionism.[75]

Depending on how we interpret the expression 'higher-level cognitive processing' in the preceding argument, we can generate stronger and weaker versions of the argument. The strong version of the argument from simple minds results when we interpret the claim that animals have no higher-level cognitive processes as the claim that animals have no introspective apparatus (be it an inner eye or a generator of higher-order beliefs). If animals lack such machinery, and thus lack the capacity to introspect, then introspectionism predicts that animals are unconscious. Therefore, the existence of animal consciousness falsifies introspectionism understood as the hierarchical theory of consciousness.[76]

We get the weaker version of the argument from simple minds, if we take the expression 'higher-level cognitive processing' to refer to the animal's capacity to form qualitative beliefs. If animals can experience qualitative phenomena while being incapable of forming the relevant qualitative beliefs the belief thesis must be given up. For then all the qualia that animals have are detached qualia. And, as was argued above, an objection to the belief thesis is also, if only derivatively, an objection to the thesis of introspectionism.

The strong version of the argument from simple minds is more deserving of our attention than the weak version. First, it makes a much bolder claim and actually introduces a new and more direct attack on introspectionism. Second, the weak version of the argument has a certain air of frivolity about it. It is doubtful that champions of the belief thesis ever intended to extend their thesis beyond the human case. Therefore, it is unlikely that anybody would defend introspectionism as a general theory of consciousness (i.e., for human and nonhuman animals alike) on the basis of the belief thesis. It seems, then, that the weak version of the argument from simple minds is directed at a straw man. In the remainder of this section I will, therefore, focus on the strong version of the argument from simple minds.[77]

The simplicity of the argument from simple minds is offset by the difficulty of assessing its soundness. Both premises can be, and have been, challenged. The first premise—that animals have conscious states—has been denied for a

number of reasons. The Cartesian worry that they lack a soul and are, therefore, insentient falls outside the scope of our discussion. But the doubt that animals have conscious sensations may stem from a different source. One's worry might, for example, be fueled by a tacit commitment to the very theory of consciousness that I have been criticizing all along. If you think that a being can have a phenomenology only if it is capable of introspection and you further believe that animals are incapable of introspection, then you will be led to doubt that animals do enjoy qualitative experiences. Here is how Lloyd puts this point:

> Simple minds, being exempt from introspection, are unaware that they have sensations. Thus if one feels an impulse to deny sensation to the simple, I suspect it traces to the tendency to confuse having a sensation with being (introspectively) aware that one is having a sensation. The latter may be our speciality; but the former is widespread in the animal world. (Lloyd 1989: 199)

I agree with Lloyd that this 'reason' for doubting animal consciousness is no more than a confusion. To my mind neither the lack of a Cartesian soul nor that of an introspective capacity raise doubts about animal consciousness.[78]

Defending the second premise of the argument from animal consciousness is a more daunting task. The case for attributing introspective capacities to animals has been made rather forcefully in recent times. Nicholas Humphrey, for example, has argued that animals who participate in sophisticated social interactions must have developed extensive introspective capacities. Here is the gist of his thesis that humans and other animals living in complex social groups are *natural psychologists:*

> A revolutionary advance in the evolution of mind occurred when, for certain social animals, a new set of heuristic principles was devised to cope with the pressing need to model a special section of reality—the reality comprised by the behavior of other kindred animals. The trick which nature came up with was *introspection;* it proved possible for an individual to develop a model of the behavior of others by reasoning by analogy from his own case, the facts of his own case being revealed to him through examination of the contents of consciousness. (Humphrey 1980: 30)

These considerations lead Humphrey to attribute introspective capacities to, for example, "wolves, chimpanzees and elephants" and to withhold such attribution from, for example, "frogs and snails and codfish." (Humphrey 1980: 35) If correct, this conclusion would deal a severe blow to the argument from animal consciousness. For it seems that the animals that Humphrey endows with introspection are precisely the ones to which we would have felt most comfortable attributing a phenomenology. Those animals that are not natural psychologists

also appear less likely to be judged to have qualitative conscious experiences. But it is not obvious that the argument from animal consciousness must be given up if Humphrey's thesis is correct. It does not seem absurd to assume that fishes, for example, enjoy a phenomenology. If this were so, it would still suffice to show that there can exist qualia that are being had without being introspected, even if we accept Humphrey's thesis of higher social animals as natural psychologists.

But others have taken Humphrey's considerations to have much more far reaching implications. Donald Griffin, for one, boldly includes social insects, bees, and ants, for example, into the family of natural psychologists. (See Griffin (1984: 186-191)) If Griffin is right I see no way of upholding the argument from animal consciousness. The set of animals to which I would be prepared to attribute a phenomenology seems to be a small subset of Griffin's introspective animals. Thus, the plausibility of the claim that animals afford examples of beings that sport unintrospected qualitative states dwindles to nothing.

Of course we could also choose to view Griffin's incontinent proliferation of self-reflective animals as a *reductio* of the thesis of natural psychologists. Thomas Reid may have overstated the case when he wrote that the power of reflection upon the operations of one's own mind is so difficult that "most men seem incapable of acquiring it in any considerable degree." (Reid 1785: 42) But surely Griffin is off by just as much when he praises the introspective powers of ants.

I do not want to place an inordinate weight on the argument from animal consciousness. I think it has a certain plausibility. But my confidence in the existence of conscious but nonintrospective animals is not unshakable. And I hope that my opponent will, in turn, refrain from claiming to know that there are no conscious but nonintrospective animals. The facts are not in and for the time being we have to be content with speculation. Neither one of us should sink so low as to rely on the kind of trick employed by the first philosopher in Chuang-Tze's story of the two philosophers on the bridge. (See Russell (1954: 180)) The first says: "See how the little fishes are darting about. Therein consists the pleasure of fishes." The second replies: "How do you, not being a fish, know wherein consists the pleasure of the fishes?" To which the first retorts: "How do you, not being I, know that I do not know wherein consists the pleasure of fishes?"

5.10 Summary

In this chapter I attempted to understand the lure of introspectionism. I argued that the view that introspection sheds light on consciousness springs primarily from one's failure to distinguish one's being conscious from one's being con-

scious that one is conscious. Our knowledge of our consciousness is best explained by appeal to introspection; what it is to be conscious is not. If these two questions are not properly distinguished the correct answer to the first problem will be mistaken for an acceptable solution of the second problem.

But the confusion of consciousness with the consciousness of consciousness grows out of a deep insight embodied in introspectionism. It is the insight that consciousness is always consciousness *for someone*—that it has built into it a direction upon an apprehending subject. Introspectionism conceptualizes this truth wrongly by cashing it out in introspective terms. This, I suggested, is a mistake that lends a spurious credibility to the idea that introspection holds the key to consciousness. Our goal must be to preserve this insight but to do it in a way that does not rely on the hierarchical apparatus employed by the various versions of introspectionism. The subject involvement of consciousness must be captured without relying on higher-order representations (HOR), no matter whether these representations are conceived of as higher-order thoughts (HOT) or higher-order perceptions (HOP).

The bulk of the chapter was devoted to an analysis of the origins of the level confusion underlying introspectionism. I suggested that three epistemic theses—collectively referred to as the thesis of constant cognition—are responsible for the intuitive appeal of introspectionism. But introspectionism, now quite unsupported by a defunct epistemology, continues to exercise its attraction propelled, it would seem, by the powerful forces of cognitive inertia.

The epistemic argument for introspectionism formed the heart of the discussion. If one holds that (i) qualitative states are invariably cognized by means of an accompanying qualitative belief and one also holds that (ii) a qualitative state must be introspected in order to become the object of an accompanying belief, then one has forged a connection between qualitative states and introspection. This connection suggests the hypothesis that being conscious simply consists in a mental state's being the object of an act of introspection. Based on Russellian considerations about acquaintance, I argued that a fairly strong case can be made for (ii). But the bulk of the discussion is devoted to (i).

The belief thesis—the thesis that our conscious mental states are necessarily accompanied by a belief in their existence—was singled out as the most conservative and therefore the most defensible formulation of the claim that we enjoy a special form of cognitive access to our qualitative mental states. The truth of the belief thesis strongly supports the introspectionist position. For in order to form the requisite qualitative beliefs, the subject must be able to become introspectively aware of his qualitative states. Thus, the belief thesis makes plausible the claim that explaining consciousness is a matter of explaining introspection. In criticizing the belief thesis I argued that having a qualitative belief is neither suf-

ficient (the thesis of detached qualitative beliefs) nor necessary (the thesis of detached qualia) for having a particular qualitative, conscious state. Consideration of cases of mistaken qualitative beliefs (like the ones that were considered in the discussion of the special-knowledge thesis) suffice to refute the sufficiency claim. The appeal to phenomenology and the discussion of animal consciousness showed that having a qualitative belief is not necessary for one's being in a certain qualitative state. With the breakdown of the belief thesis introspectionism is robbed of its most crucial support. Though not refuted outright, it now presents itself as a mere unsupported possibility.

The strong version of the argument from simple minds took the attack on introspectionism one step further by taking a direct (though rather speculative) shot at introspectionism. If sound, the strong version of the argument from simple minds shows that introspection is not a necessary condition for qualitative consciousness.

For the most part the argument of this chapter attacked introspectionism indirectly. I tried to show that the main source of introspectionism's plausibility is dry. In the absence of further support for introspectionism it therefore appears reasonable to treat it as an unsupported hypothesis. In other words, there appears to be no good reason for thinking that consciousness can be explained by appeal to introspection. Far from being obvious, the connection between consciousness and introspection is quite unsupported.

In the next chapter I present a less indirect critique of introspectionism. Focusing on Pollock's version of the doctrine—as it is embodied in Oscar, the artificial mind Pollock is building—I argue that we have good reasons to reject introspectionism.

CHAPTER SIX

Oscar, the Unconscious Introspector: A Case Study

In the previous chapter I tried to drive a wedge between consciousness and introspection. My main goal was to rob introspectionism of its aura of plausibility. I argued that the epistemic views that made introspectionism seem plausible in the first place should be discarded. If introspectionism continues to seem plausible, this is so not because of its intrinsic merits but because we have failed to realign our intuitions with our new convictions concerning the epistemic access to our conscious mental states.

In this chapter I will continue to pursue the goal of discrediting introspectionism. But I will proceed in a different manner. Instead of talking about introspectionism in general, I will focus on Pollock's version of a hierarchical theory of consciousness. And instead of attacking introspectionism in a roundabout way through a critique of the source that nourishes it, I will try to show that Pollock's introspectionist construction does not suffice to serve up consciousness. Nor do I discern a reason for believing that the Pollockian architectonic is necessary for consciousness.

The critique of Pollock's introspectionism cannot take the detour through the critique of the belief thesis because Pollock does not base his hierarchical theory of consciousness on the belief thesis (or on any thesis like it). Remember that Pollock was one of our star witnesses for the phenomenological case against the belief thesis. Remember also that Pollock supplied some of the arguments against the special-knowledge thesis. Pollock's reasons for going the introspectionist route must therefore stem from some other source. In the last chapter I listed three reasons that might (mis)lead one to perceive a connection between introspection and consciousness: the definitional equation of the two; the thesis that introspection best explains consciousness; and the level-confusion between consciousness and consciousness of consciousness. I think that Pollock's version of introspectionism is best viewed as motivated by the best explanation strategy. Pollock develops a detailed model of our cognitive functioning and points out that the having of qualia becomes entirely unmysterious if we identify it with the output of certain introspective sensors that are part of our cognitive architecture. I will argue that we should resist this identification because we have

good reasons for believing that a being instantiating Pollock's hierarchical theory of consciousness can, and does, lack all qualitative consciousness. I conclude that introspectionism—as exemplified by Pollock's Oscar—fails as a theory of consciousness.

6.1 Pollock's Oscar as Imitation Man

In his book *How to Build a Person* Pollock (1989) really does endeavor to tell us how to build a person—surprisingly enough the catchy title is accurate. Pollock is currently engaged in the project of building an artificial person—Oscar—and his book is a sketch of the philosophical underpinnings of this venture.[79] Oscar's creator maintains that once Oscar is complete he "will experience qualia, will be self-conscious, will have desires, fears, intentions, and a full range of mental states." (Pollock 1989: ix-x) I will focus on what seems to me to be the most problematic of these claims, viz., that Oscar will experience qualia. I argue that we have not been given sufficient reasons to believe this bold claim. I doubt that Oscar will enjoy qualitative conscious phenomena and maintain that it will be like nothing to be Oscar. Borrowing a term from Keith Campbell, I can put my worry by saying that Oscar will be an imitation man—a being very much like us except that "instead of feeling a pain when he burns his finger…he has no locatable sensations at all. He just spontaneously gains a new belief, it just "pops into his head" that he has burned his finger…" (Campbell 1984: 100) The charge I will level against Oscar is not exclusively concerned with pain qualia. I will ask you to consider the possibility that Oscar is an imitation man through and through, that all his perceptual accomplishments proceed unmediated through, and unaccompanied by, any phenomenology at all. But before we can consider these possibilities, we will have to sketch a fairly detailed account of Oscar's philosophical underpinnings. I now turn to this task.

6.2 Rational Functionalism and Consciousness

6.2.1 Linking Rationality and Consciousness

Consciousness poses a difficult problem for functionalism. The attempts do deal with this problem span the whole range from Dennett's eliminativist advice that it is best "to declare that there simply are no qualia at all" (Dennett 1988: 44) to Pollock's view that a functionalist account of rationality *cannot* renounce the task

of providing an account of a full blown notion of consciousness including its qualitative glow.

Pollock bases this view on two claims. First, he holds that a functional theory of the mind is, essentially, a theory that describes the rational relations obtaining between mental states; i.e., a functional theory of the mind is a theory of the rational architecture of the human mind. And second, he argues that some common forms of reasoning require that certain mental states of the reasoner be conscious. Thus, a functionalist theory of the mind, i.e. a functional theory of the human rational architecture, will have to account for conscious mental states. By specifying the role that consciousness plays in our mental economy Pollock's rational functionalism overcomes a major problem that previous attempts to provide a functional account of consciousness had to face. As long as it was not understood how consciousness interacts with our other mental states, as long as it seemed that consciousness did not play any relevant part in our mental life, consciousness was not amenable to a functional treatment. For, simply put, one cannot provide a functional account of something that has no function.

In the course of addressing the question whether there could be robots who behaved just like people but experienced nothing, Nagel states an intriguing hypothesis but does not stop to work it out: "Perhaps there could not actually be such robots. Perhaps anything complex enough to behave like a person would have experiences." (Nagel 1974: 167) Pollock's theory confirms Nagel's conjecture: a robot who behaves like a human being in all respects, cannot but have a full blown conscious mental life. If a robot is to behave like us, it will have to behave rationally in just the way we do. And in order to mimic our rationality, the robot's behavior will have to be governed by the human rational architecture. And possession of the human rational architecture entails having all the mental characteristics that Nagel could wish for: "A system modeling human rationality in the appropriate sense will experience qualia, will be self-conscious, will have desires, fears, intentions, and a full range of mental states." (Pollock 1989: ix-x)

Discussing a related issue, Haugeland coined the phrase: "If you take care of the syntax, the semantics will take care of itself." (Haugeland 1981: 23) Making the necessary substitutions, we can put Pollock's point as follows: *if you take care of the rationality, the feel will take care of itself.*

6.2.2 The Role of Consciousness in the Human Rational Architecture

Why should the possession of the human rational architecture entail the possession of consciousness? The aspects of our rational architecture that feature in Pollock's explanation of this entailment all share one feature: they presuppose a certain amount of self-reflexiveness on our part. It is, according to Pollock, this

self-reflexive structure of our reasoning that introduces the warm phenomenal glow of consciousness into the world.

Pollock describes three forms of reasoning that require us to be aware and to keep track of at least some of our mental states. (i) We exercise foresight when, for example, we avoid imminent pain; to do that we must be able to keep track of prior pain sensations. (ii) In order to form manageable generalizations about our environment, we must on occasion overrule the dictates of our senses. To do that, we must be able to keep a track record of the successes and failures of our sensory apparatus. Based upon this record, we form beliefs about the reliability of our senses. And these beliefs will then guide us in rejecting or accepting the testimony of our senses. (iii) We retract conclusions we reached, when we find that further reasoning leads to conclusions that undermine our original reasoning process—i.e. we reason defeasibly. To do that, we must be able to keep track of the reasoning process that led us to our original conclusion. To keep track of our mental states (of our pains, of the deliverances of our perceptual sensors, and of our thoughts), we rely on our capacity to introspect (some of) our mental states. And this introspective access to our own mental states is all of a piece with the enjoyment of qualitative phenomena. Introspection "endows" (Pollock 1989: 51) the occurrences to which it is directed with qualitative properties. In introspection our mental states present themselves to us in the guise of the phenomenal qualities that pervade our mental life. This concludes the rough sketch of Pollock's view that possessing the rational architecture we do leads us to enjoy the characteristic kind of consciousness that makes it like something to be us.

6.3 How to Build Consciousness into Oscar

Let us assume that we encounter no problem in providing Oscar with deductive and inductive reasoning capacities, with a goal-structure, and with sensors that allow him to respond to the environmental conditions he encounters. As presently described, we may think of Oscar as a machine conceived in a truly behaviorist spirit. Karl Lashley's dictum *no activity of the mind is ever conscious* is literally true of him. On this developmental level Oscar will not be able to engage in any of the sophisticated forms of reasoning listed above because he is not able to keep track of any of his mental occurrences. Pollock suggests that providing Oscar with the necessary 'self-awareness' is simply a matter of installing a second set of sensors. Some of these second-order sensors detect activity in Oscar's first-order sensors; others monitor occurrences that we may call Oscar's thoughts. We will call the first-order sensors 'perceptual sensors'; second-order sensors we call 'introspective sensors.' Thus Oscar turns into an in-

trospective machine. Using his introspective sensors, Oscar can detect and track his thoughts and his sensations. Having achieved this degree of self awareness, Oscar is now in the position to find out under which circumstances his pain sensors usually fire; thus he will be able to foresee and avoid potentially painful situations. His sensor-sensors will enable him to determine under which conditions his perceptual sensors are prone to mislead him; thus he will no longer try to shake the hand of the nice robot in the mirror. And insofar as he employs his thought-sensors and remembers the steps that took him to his conclusion, he will later be able to retract his conclusion in the light of potential defeaters for his reasoning.

In implementing this second layer of Oscar's rational architecture, we have not only given Oscar the capacity to reason in sophisticated ways, we have also given him qualitative consciousness. For "having a qualitative feel is just a reflection of the operation of the introspective sensors." (Pollock 1989: 14) Put a little more precisely, one can say that the "qualitative feel is the output of the introspective sensors." (Pollock 1989: 14) Applied to the specific example of pain this yields the result that "the "feel" of pain is the output of the pain sensor sensor." (Pollock 1989: 11)

Pollock's account of qualitative consciousness bears the familiar features we encountered when discussing HOR accounts of the having-problem in Chapter Four. It also raises the same questions. One wonders how this simple story could possibly suffice to switch on Oscar's inner light. Pollock feels the intuitive power of this objection which he states simply and forcefully: "One may protests that the "feel" of the qualitative feel is being left out of the picture." (Pollock 1989: 14) Turning to the example of pain again, he articulates the worry like this: "The qualitative feel of pain is too prominently a part of pain, and any physical account *seems* to leave it out." (Pollock 1989: 14) In a footnote he clarifies this objection by employing Nagel's 'what it's like' locution. What seems to be left out is that "an essential feature of pain is that there is something that it is like to be in pain." (Pollock 1989: 14) It seems, then, that the objection Pollock is considering is concerned with what I have been calling qualitative consciousness. So we might briefly state the objection by saying that it appears that Oscar lacks qualitative consciousness.

In good analytic style Pollock replies by drawing a distinction—tacitly implying, of course, that those who worry about the feel are confused on this point. He writes:

> In this connection we can usefully distinguish between *having* the qualitative feel and *experiencing* the qualitative feel. To have the feel is for one's first-order introspective sensors to be operating, but to experience the feel is to attend to the feel itself, and that consists of the operation of the second-order

> introspective sensors sensing the operation of the first-order introspective sensors. (Pollock 1989: 14)

We have here a three-leveled structure: three levels of sensors and corresponding to the activity of the sensors on each of these levels we get three distinct outputs. Consider what happens in Oscar when he looks at a ripe tomato:

1. A perceptual sensor operates and produces a red sensation.[80]
2. A first-order introspective sensor operates on the output of the perceptual sensor and produces the having of a qualitative feel of redness.
3. A second-order introspective sensor operates on the output of the first-order introspective sensor and produces the experiencing of the qualitative feel of redness.

At which level does consciousness set in? At which level of this multi-layered process does it start to be like something to be the one who undergoes these processes? I think that Pollock holds that nothing less than level (3) will guarantee a fully conscious perceptual experience. Consider level (1) in isolation: the red sensation, the output of the perceptual sensor, does not bring any qualitative aspects into play. When it occurs, unregistered by any higher-level sensors, it neither has, nor is it experienced as having, any phenomenal qualities whatsoever. In other words, it is like nothing to undergo such an automatic reaction of your perceptual sensor. A certain mental state is a red sensation solely in virtue of the role it plays in a particular rational architecture; that it have, or that its owner take it to have, a certain phenomenal property is not required. Adding level (2) to the operation of level (1) changes the phenomenon: being registered by a first level introspective sensor endows the red sensation with the phenomenal property of redness. Oscar may now be said to have—but not experience—a red quale. Only on level (3) does Oscar get to experience the red qualitative feel. Level (3) is supposed to answer those who protested that "the "feel" of the qualitative feel is being left out of the picture." Since Pollock also characterized this 'feel of the qualitative feel' with Nagel's 'what it's like' locution, one gets the impression that only on level (3) will it start to be like something to be Oscar looking at the ripe tomato.

This seems to me the most likely but not the only possible interpretation of Pollock's theory. The alternative I have in mind is this. Using my own terminology the operation of Oscar's three levels of sensors might be described as follows (when confronted with the ripe tomato):

1. A perceptual sensor operates and produces a mental state in Oscar that exemplifies a red celluloid quale.
2. A first-order introspective sensor operates on the output of the perceptual sensor and lets Oscar have the red quale.

3. A second-order introspective sensor operates on the output of the first-order introspective sensor and makes Oscar fully introspectively conscious of the red quale.

On this rendition it becomes like something to be Oscar on level (2). Level three would merely add the focused awareness of one's qualitative consciousness that phenomenologists presumably enjoy (see Section 4.2.2.1). On level (3) Oscar would become 'conscious of his consciousness.' Interpreting Pollock in this way aligns his view more closely with those of the other HOR theorists. I am inclined to resist this interpretation because it does not do full justice to some of Pollock's claims about consciousness. In the end not much turns on which interpretation one chooses. For the points I try to make against Pollock's account do not seem to depend on the question whether consciousness is claimed to enter the world on level (2) or level (3).

6.4 Doubting that Oscar Is Conscious

In the preceding section I tried to present Pollock's account of why we should admit Oscar into the charmed circle. Oscar possesses the human rational architecture and must, therefore, be conscious. Now I will present my reasons for doubting that Oscar is conscious. I will maintain that we can understand how Oscar reasons, that is, that we can understand the operation of the multi-layered system of perceptual and introspective sensors that supports his reasoning, without having to attribute a subjective felt quality to any of the events that transpire during the process of his reasoning. Therefore, the claim that Oscar must be conscious is untenable. Further considerations make it reasonable to accept the hypothesis that all the processes that underwrite Oscar's reasoning do in fact take place in perfect inner darkness. I conclude that it is reasonable to assume that Oscar is unconscious.

6.4.1 Perception without Sensation

We can get started by looking at a way of thinking about perception that is officially endorsed by D.M. Armstrong and that also seems to be at work in Pollock's discussion of the nature of consciousness.[81] Armstrong writes:

> Eccentric cases apart, perception, considered as a mental event, is the acquiring of information or misinformation about our environment. It is not an "acquaintance" with objects, or a "searchlight" that makes contact with them, but is simply the getting of beliefs. Exactly the same must be said of introspection. It is the getting of information or misinformation about the current state of our mind. (Armstrong 1968: 326)

In another place Armstrong emphasizes the link between perception and selective behavior: "In general, we can think of perceptions as inner states or events apt for the production of certain sorts of selective behavior towards our environment." (Armstrong 1966: 13) As one critic notes, the most striking feature of Armstrong's characterization of perception surely is his "disregard for *sentience*." (Kneale 1969: 299)[82] No mention is made of sense organs, of sensations that are merely had or of sensations that are experienced. Nothing in these characterizations of perception indicates that the perceiver experiences anything during the act of perception. Given this notion of perception, it may well be like nothing to perceive one's surroundings.[83] This same line of thought is pushed a bit further by Paul Churchland. The passage I am going to quote clearly expresses the consideration that leads me to doubt that it is like anything to be Oscar:

> Sensations are just *causal* middle-men in the process of perception, and one kind will serve as well as another so long as it enjoys the right causal connections. So far then, in principle they might even be *dispensed* with, so far as the business of learning and theorizing about the world is concerned. As long as there remain systematic causal connections between kinds of states of affairs and kinds of singular judgments, the evaluation of theories can continue to take place. (Churchland 1979: 15)

I take it that Churchland (unlike Pollock) uses the term 'sensation' to refer to the experienced feel of a perceptual episode. I will adopt his use of the term because it facilitates the discussion of the problem. What he is saying, then, is that it is possible to perceive X without accompanying sensation of X. You can, as it were, short-circuit the perceptual process, thereby rendering the qualitative element of the process superfluous. In countenancing this possibility, Churchland is not denying the obvious fact that *our* perceptions are accompanied by sensations. As a matter of fact, *we* do perceive our surroundings by having sensations. We can pay more or less attention to our sensations and mostly we do pay little if any attention to them; this results in our 'looking through' our sensations. But the transparency of our sensations must not be mistaken for their nonexistence: that we do have sensations follows from the fact that we do perceive our surroundings. Churchland urges that while this relation between perception and sensation is true for us, it need not be true in all cases. In urging us to take seriously the possibility of perception without sensation Churchland advocates 'conscious inessentialism' (see page 18) with regard to perception.[84]

We now know that the possibility that was adumbrated in Armstrong's characterization of perception and explicitly pointed out by Churchland is actualized in those individuals who suffer from blindsight. Blindsight "delivers true belief in—capacity to point to—the position of visual stimuli without either assent [i.e.

conscious belief] or visual sensation."[85] More generally, blindsight is the capacity to perform visual discriminations in the absence of any visual sensations on the part of the subject. The blindsighted subject who performs a visual discrimination task takes herself to see nothing. To perform a task of visual discrimination feels like nothing to her. Blindsight is precisely the kind of phenomenon that Churchland envisioned, viz., perception (understood as the capacity to behave selectively towards your environment) without sensation (understood as experiencing a phenomenal quality corresponding to the type of perception at hand).

6.4.2 The Lessons of Blindsight

The existence of blindsighted persons establishes as a fact what philosophers had countenanced as a mere possibility: that there can be perception without sensation. The existence of this phenomenon is the pillar upon which my skepticism concerning Oscar's consciousness rests. But before I can present 'the argument from blindsight,' I must say something about Pollock's treatment of blindsight. For, unlike me, he thinks that his theory can derive some support from blindsight: he takes its capacity to explain this baffling phenomenon to be a showcase of his theory's power.

I will, therefore, begin by briefly describing Pollock's explanation of blindsight. I will then present some empirical evidence (concerning the capacities of blindsighted persons) that is inconsistent with Pollock's explanation of blindsight. I will argue that Pollock gets caught in a dilemma from which I do not know how to free him. Finally I will present 'the argument from blindsight' from which I conclude that consideration of the phenomenon of blindsight affords us a perspective from which it appears that Oscar might be entirely unconscious.

6.4.3 Pollock's Analysis of Blindsight

We already encountered Pollock's characterization of blindsight when reflecting on the nature of the conscious state of Armstrong's absentminded driver (see Section 4.2.2.1, page 63). Pollock holds that a blindsighted person neither has nor experiences qualia. A blindsighted person is one whose visual sensors are operational but whose introspective sensors don't work. Such a person is capable of reacting selectively towards her environment. But since her introspective sensors are not working, her perceptual achievement is not accompanied by any qualitative state or experience thereof.

Given Pollock's account of the human rational architecture, his explanation of blindsight does look very plausible indeed. I do, however, want to point out a consequence of this account that will soon prove to be troublesome. On Pol-

lock's view, the three tiered sensor structure of a normal individual is reduced to a single layer of perceptual sensors in the blindsighted individual. The blindsighted individual will therefore not only miss out on the enjoyment of the qualitative aspect of visual perception; the blindsighted individual will also be incapable of performing any task that requires her to monitor or keep track of her visual sensations. I will now argue that this consequence of Pollock's account is incompatible with certain observations about the capacities of blindsighted subjects.

6.4.4 Introspection and Learning

The first and strongest pressure to fit Oscar with introspective sensors arose from the following consideration: Oscar's pain sensors allow him only to *react* to damaging situations, by fight or flight, for example. But by the time Oscar actually encounters the damaging stimulus it may already be too late. Oscar needs to be able to anticipate and avoid such situations. In order to make him capable of exercising foresight in this manner, Oscar had to be fitted with introspective sensors, with pain-sensor sensors. This is how Pollock describes the difference between Oscar before he has any introspective sensors (Oscar I) and Oscar after pain-sensor sensors are installed (Oscar II):

> It is illuminating to note that the difference between Oscar I and Oscar II is roughly the difference between an amoeba and a worm. Amoebas only respond to pain …; worms can learn to avoid it. The learning powers of worms are pretty crude, proceeding entirely by simple forms of conditioning, but we have said nothing about Oscar that requires him to have greater learning powers. (Pollock 1989: 2)

I want to focus on the notions of *learning* and *conditionability* that Pollock employs to mark the difference between Oscar I and Oscar II. Before Oscar had any introspective sensors, he lacked even the most rudimentary capacities to learn. After the installation of introspective sensors, Oscar did have the capacity to be conditioned. Pollock's discussion does suggest that the possession of introspective sensors is a necessary condition for a system to be capable of being conditioned. Had Oscar not been given pain-sensor sensors, he could not have learned how to foresee and avoid damaging stimuli. The claim (which I take to be Pollock's claim) that the possession of introspective sensors is a necessary condition for a subjects capacity to learn via simple conditioning is one that I will rely on in the following discussion.

6.4.5 The Capacities of Blindsighted Subjects

In this section I want to describe briefly some of the capacities of blindsighted monkeys and men. Helen was a rhesus monkey whose visual striate cortex had been surgically removed. This is what happened after the operation:

> Immediately after the operation she appeared to be quite blind, and during the next year she showed little sign of any vision. But then watching her closely, we [Nicholas Humphrey and L. Weiskrantz] saw that occasionally she turned her eyes to look at moving objects. I tried to encourage her in this unexpected behavior by giving her bits of food whenever she made a directed eye-movement and with a few days her performance dramatically improved—not only did her eye-movements become more and more reliable but she was soon persuaded not just to look but to reach out with her hand for the moving object. (Humphrey 1972: 682)

Eventually Helen's "recovery seemed so complete that an innocent observer would have noticed very little wrong with the way she analyzed the visual world." (Humphrey 1972: 37-38; for more on Helen see Humphrey (1974); Humphrey and Weiskrantz (1967)) This account leaves no doubt about the fact that Helen was capable of being operantly conditioned.

Similar, if much less dramatic, results have been achieved with human subjects. Stimulated by Helen's success and maybe also by some lingering doubt whether Helen was really blindsighted, researchers have undertaken to investigate human blindsighted subjects, in particular the question whether such subjects could learn to 'see better.' Blindsighted persons were trained to focus their eyes on targets (usually bright spots of light flashed onto a neutral background) that were presented very briefly in the blind region of their visual field. Despite the fact that these subjects took themselves to have seen nothing, their "ability to detect and localize stimuli improved markedly after the systematic training of visually evoked saccadic eye movements." (Zihl 1980: 71-77) Weiskrantz, arguably the leading authority on blindsight, describes these results in terms of response shaping (a form of operant conditioning). Speaking about results that were obtained with monkeys and with human subjects, he says that "the appearance of blind-sight may depend heavily…on practice and shaping up." (Weiskrantz 1980: 380-382)

This cursory report on work done with blindsighted subjects shows that they can be conditioned to perform better in tasks of visual discrimination. I will now argue that Pollock's account of blindsight is incompatible with this documented capacity of blindsighted subjects to learn.

6.4.6 Blindsight and Learning: A Dilemma

The existence of conditionable but blindsighted subjects poses a problem for Pollock's account of the mind. For his account has the consequence that a subject (with the right kind of rational architecture) cannot be blindsighted, if it has a functional organization rich enough to allow for conditioning. This is so because, on Pollock's view, the mechanism underlying both capacities is the same: the subject becomes conditionable in virtue of having and exercising introspective sensors. But a subject that has and exercises its introspective sensors will have and experience qualia and will, therefore, not suffer from blindsight. Thus, there cannot be a subject that is both conditionable and blindsighted.

In order to restore consistency, Pollock will have to accept one of the following alternatives, both of which are damaging to his account of the mind and consciousness. Either it is possible for subjects to perform certain sophisticated cognitive tasks without the involvement of any self-monitoring on the part of the subject, or it is possible for a subject to monitor some of its own cognitive processes without there arising any qualitative phenomena in the subject.

If Pollock chooses the first horn of this dilemma, then the notion of a conditionable but blindsighted subject no longer poses a problem. But the price he would have to pay is high. The acknowledgment of the existence of sophisticated cognitive processes that can proceed without the involvement of any higher-order sensors would amount to nothing less than the abandonment of the project Pollock set out to complete. The idea that the explanation of sophisticated cognitive processes is to be carried out by appeal to a hierarchy of interacting sensors is at the heart of his theories of the mind and of consciousness. If this part of the theory were abandoned, Pollock would have to start anew.

The second horn of the dilemma is no more attractive than the first. If the possibility of a conditionable but blindsighted subject can only be accommodated by giving up the thesis that qualitative feel just is the output of the introspective sensors, then Pollock's theory would suffer a severe setback. To give up this thesis is to give up one of the central and most interesting theses of Pollock's theory of consciousness. In fact, if this claim were given up, Pollock would no longer have a theory of consciousness.

I conclude that the phenomenon of blindsight poses a problem for Pollock's theory of consciousness. His initially very plausible explanation of blindsight breaks down in the light of further evidence about the capacities of blindsighted subjects. Instead of deriving support from its purported ability to explain the puzzling phenomenon of blindsight, blindsight must be viewed a serious problem for Pollock's account of consciousness.

So far I have only argued that Pollock does not provide a satisfactory explanation of blindsight. This may strike one as a rather unimportant result since blindsight is, after all, a rather peripheral phenomenon that can comfortably await treatment at some later time. But I will now try and argue for a stronger conclusion: reflection on the nature of blindsight affords us a reason to doubt that Oscar is conscious at all.

6.4.7 The Argument from Blindsight

Pollock's own way of conceptualizing blindsight as resulting from the failure to have or exercise one's introspective sensors leads straight into the horns of a dilemma. I will therefore propose a different way of thinking about the phenomenon of blindsight. When viewed from this angle, consideration of the phenomenon of blindsight affords a reason for doubting that Oscar has or experiences any qualitative phenomena whatsoever.

Oscar has three kinds of sensors: perceptual sensors, and first- and second-order introspective sensors. What we have here are three levels of sensors layered on top of each other. A primitive Oscar model with only the first level perceptual sensors would perceive the environment without the benefit of any qualitative accompaniments. It would be a blindsighted Oscar. Adding two additional layers of sensors to Oscar's architecture was supposed to remedy this problem. First- and second-order introspective sensors were built into Oscar so that he can have and experience qualitative phenomena. What I will try to show is that adding more sensor levels need not lead to the desired result of providing Oscar with an inner life.

Using the expression 'blindsight' in a completely generic manner to refer to *all* modes of discriminatory activity without corresponding qualitative dimension, I can express my worry about Oscar thus: *Oscar might be suffering from an all out case of blindsight.* His perceptual sensors, his first-order introspective sensors, and his second-order introspective sensors might all be 'blind.' With Oscar it might be blindsight all the way up. Under these circumstances a subjective feel would not appear on any of the three tiers of Oscar's hierarchy of sensors. And if not a single one of Oscar's three perceptual and introspective sensor layers spawns qualia, then surely Oscar as a whole is devoid of qualia too. Note that this impairment would in no way interfere with Oscar's spotless performance as a reasoner. He could discriminate and behave selectively towards his surroundings, towards the outputs of his perceptual sensors, and towards the output of his first-order introspective sensors. A blindsighted Oscar could keep track of the activity in his pain sensors and so learn to exercise foresight in avoiding future painful situations. He can detect the activation of his red sensor

which will then allow him to discover its partial unreliability. He can carry out all these self-monitoring tasks and all the forms of sophisticated reasoning for which they are a prerequisite, the only difference being that they will not be accompanied by a phenomenal glow. *A thoroughly blindsighted Oscar is unconscious; he does not experience anything in any way.*

6.4.8 An Objection to the Argument from Blindsight

It must be granted that each one of Oscar's three layers of sensors, when taken separately, could operate without producing qualia. But the claim was that Oscar, not any one of his functional components in isolation, is capable of having and experiencing qualitative phenomena. From the fact that none of his isolated components produces qualia it does not follow that Oscar does not have or experience qualia. To reason in this manner from the properties of the components to the properties of the whole that these components make up is to commit the fallacy of composition.

6.4.9 Reply to the Objection

This objection does lay to rest all attempts to construe the argument from blindsight as a deductively valid argument. But this is not the end of the story. The argument from blindsight cannot serve as a proof that Oscar is unconscious. That much the objection has established. But the argument can, nevertheless, induce something akin to a *gestalt* switch. Instead of asking: How could Oscar be monitoring his internal states without his having and experiencing qualia? one may now feel the need to ask: How could Oscar's self-monitoring give rise to his having and experiencing qualia? Oscar performs discriminatory tasks on three levels: with his perceptual sensors he discriminates between external stimuli; with his first-order introspective sensors he discriminates between the outputs of his perceptual sensors; and with his second-order sensors he discriminates between the outputs of his perceptual and his first-order introspective sensors. All parties in the dispute agree that each one of these discriminatory tasks can be executed unconsciously. In this situation our question must be: How are we supposed to understand the claim that the joint execution of these acts of discrimination amounts to Oscar's having and experiencing qualia? As long as the manner in which the orchestrated employment of Oscar's sensor hierarchy is supposed to give rise to qualitative phenomena remains unintelligible, we are not even in a position to sensibly discuss the question whether we have been given good reasons for believing that Oscar has and experiences qualia in virtue of the functioning of his three-tiered structure of sensors and sensor sensors. If this question does now strike you as urgent, then the argument from

blindsight has done its job, viz., to shift the burden of proof to our opponent. And as long as an answer to this question is not forthcoming, we will be well advised to remain skeptical about Oscar's consciousness.

Pollock approaches this issue from a different angle. On his view the claim that Oscar's self-monitoring activity gives rise to his having and experiencing qualia is a specific instance of a more general conclusion derived from an argument that presents token physicalism as the best explanation of two known facts. These facts are (i) that the mental and the physical are involved in a single causal nexus; and (ii) that the physical world is causally closed. These two facts are best explained by assuming that mental events are physical events. (See Pollock (1989: 18-19)) The only remaining problem is to isolate the particular physical events with which qualitative states can plausibly be identified. This is how Pollock arrives at the more specific claim that "qualitative feel is the output of the introspective sensors." (Pollock 1989: 14) Spelled out in more detail this claim reads as follows:

> If we were building...an intelligent machine and endowing it with a cognitive architecture like ours, we would have to provide it with the ability to introspect the operation of its perceptual sensors. Introspection would be sensitive to the values of various physical parameters of the output of the perceptual sensors. The physical parameters to which introspection is sensitive would comprise the introspective dimensions along which the machine categorized its sense experience, and two mental occurrences in the machine would be of the same qualitative type just in case introspection could not distinguish between them along the relevant introspective dimensions. This, I take it, provides a perfectly good physical understanding of our employment of qualitative states. There is nothing mysterious or nonphysical about such states. (Pollock 1989: 82)

I think that this is what we should say if we knew on independent grounds that Oscar had and experienced qualitative states. If we had this knowledge, then the problem would be merely one of finding out which of his physical/functional states could plausibly be identified with his qualitative states. But this independent knowledge we do not have. All we do have is a description of Oscar's rational architecture. And we can accept this description in its entirety and still remain completely unconvinced that Oscar has or experiences any genuine qualitative states whatsoever. One who took this position would not be saying anything incoherent or absurd; nor can he be accused of wanton skepticism. As long as we are not offered an account of how the integrated activity of Oscar's sensors can serve up the qualitative glow of consciousness, the doubt that Oscar is conscious remains a reasonable one. If one goes ahead and identifies qualitative feels with the outputs of introspective sensors without bothering to explain

how this identification might be understood, then one will have to put up with the complaint that one's theory has what Keith Gunderson has so aptly labeled "that *representation-by-fiat-only* ring" to it.[86]

To my mind, these considerations show that the worry that Oscar suffers from an all out case of blindsight (and is therefore unconscious) cannot be dispelled by an appeal to the fallacy of composition. I will now turn to a discussion of reasons that should incline one to take the possibility of an unconscious Oscar seriously.

6.5 Should We Take the Possibility of Oscar's Blindsight Seriously?

I have countered the claim that Oscar is conscious (or rather the claim that Oscar must be conscious and the claim that we have good reason to believe that he is conscious) by pointing out the (nomic) possibility that Oscar suffers from all out blindsight, in which case he would be unconscious. It is often held that saying 'so what' is an adequate rebuttal to an objection that consists in nothing more than the presentation of a counter-possibility. Thus, if the blindsight-objection is to have some weight, we will have to argue that the envisaged scenario is a possibility that cannot be dismissed lightly.

6.5.1 Two Arguments for Taking Oscar's Blindsight Seriously

6.5.1.1 The Argument from Parsimony

I can think of two considerations that might incline one to take the blindsight-possibility seriously. First, respect for a kind of parsimony principle encourages us to favor less costly explanations over more expensive ones. This fact, together with the fact that we steadily accumulate "evidence that a great deal of intelligent and sentient activity, perhaps the lion's share, goes on without benefit of self-conscious awareness" (Churchland 1983: 81; see also Kihlstrom (1987)) should make blindsighted Oscar more attractive to us then his conscious brother. Both model human rationality, but the blindsighted one relies on less dubious means in doing so.

6.5.1.2 The Argument from Mystery

Sticking to Chisholm's advice: "try to keep mystery to a minimum," (Chisholm 1983: 50) yields another reason to seriously consider the possibility of the blindsighted Oscar. Here is how Nagel describes the mystery we incur by ascribing consciousness to Oscar:

> If we acknowledge that a physical theory of mind must account for the subjective character of experience, we must admit that no presently available conception gives us a clue as to how this could be done. The problem is unique. If mental processes are indeed physical processes, then there is something it is like, intrinsically, to undergo certain physical processes. What it is for such a thing to be the case remains a mystery. (Nagel 1974: 176)[87]

To my mind, Nagel points to a genuine problem for every physicalist theory of mind. For even if we grant that Oscar's possessing the architecture he does makes him into a conscious being, and even if we agree with Pollock that we now have a "perfectly good physical understanding of our employment of qualitative states," (Pollock 1989: 82) one still can be puzzled by the question how a physical occurrence, which seems to have nothing in common with an experienced feel, can manifest itself as the having of a subjective feel.[88] If this consideration is more than just begging the question against physicalism, and I think it is more, then it affords us a second reason for taking the possibility of the blindsighted Oscar seriously: the blindsighted Oscar is the less mysterious Oscar.[89]

6.5.2 An Objection: The Two Arguments Prove too Much

It may seem that the argument from parsimony and the argument from mystery suffer from the same problem: they prove too much. An objector to the argument from parsimony might argue as follows: the empirical evidence (such as it is) for the conclusion that most mental processes do or could go on without conscious awareness should lead you to take seriously not only the possibility that Oscar lacks consciousness; the empirical evidence should also lead you to take seriously the possibility that *we* are not conscious. But since you resist this latter conclusion, the argument from parsimony should not persuade you to take the possibility of a blindsighted, unconscious Oscar seriously.

The objector might run a similar argument for the case of the argument from mystery: if there is a mystery about how a physical process can give rise to qualitative mental states, then it follows that Oscar's consciousness is no more mysterious than our own consciousness. For, after all, it is no less unintelligible how the processes in a biological brain should give rise to consciousness than how the processes in a silicon brain should be capable of doing that. Thus, the argument from mystery should lead you to take seriously not only the possibility that Oscar lacks consciousness; the mystery should also lead you to take seriously the possibility that *we* are not conscious. But since you resist this latter conclusion, the argument from mystery should not persuade you to take the possibility of a blindsighted, unconscious Oscar seriously.

Both of these arguments leave out a crucial fact: we know that we are conscious. Thus the hypothesis that we are not conscious is unacceptable. For nei-

ther considerations of parsimony nor the desire to reduce mystery legitimize the acceptance of an hypothesis we know to be false. We have to live with the fact that we are built in a nonparsimonious way: we might have been designed in such a way that we could operate unconsciously, but we are not designed that way. And we have to live with the fact that we are mysterious creatures: we know that our brain produces conscious states despite the fact that we don't know how this is possible. But in Oscar's case we do not know this crucial further fact; we do not know that he is conscious. And this is why considerations of parsimony and the desire to reduce mystery should lead us to take the possibility that Oscar is blindsighted and therefore unconscious seriously.

6.5.3 Two Arguments Against Taking Oscar's Blindsight Seriously

If, as I have argued, the possibility raised by the argument from blindsight must be taken seriously, then it is unreasonable to believe that Oscar is conscious. I will now discuss two considerations that seem to show that we should not take the possibility of Oscar's blindsightedness seriously.

6.5.3.1 The Argument From Analogy

Couldn't an argument from analogy establish that Oscar's inner life must be very much like my own? If so, we would not have to take the blindsight possibility seriously.

Whatever one may think of the value of analogical arguments in the context of the problem of other minds, one must surely acknowledge that it is especially problematic to employ an analogical argument to argue from our own case to Oscar's case. Take, for example, the question whether Oscar experiences pain qualia in circumstances in which he displays pain behavior. That my screams are always accompanied by experienced pain qualia does not license the inference that Oscar's screams are accompanied by experienced pain qualia. This is so because the sample upon which the inductive evidence for the co-occurrence of screams and pain experiences is based is not fair relative to the population of Oscarites—for surely we have every reason to assume that I am not a characteristic member of the Oscar family. To my mind, this argument effectively lays to rest all attempts to establish Oscar's consciousness via an argument from analogy.[90]

Much the same point has been made by Ned Block in the context of his discussion of the consciousness of homunculi heads. He argues that it is a mistake to believe that the fact that we can establish the consciousness of our conspecifics by an argument to the best explanation licenses a similar strategy when we are dealing with homunculi heads:

> [Homunculi heads] are systems designed to mimic us, but we are not designed to mimic anything...This fact forestalls any attempt to argue on the basis of an inference to the best explanation for the qualia of homunculi-heads. The best explanation of the homunculi-head's screams and winces is not their pains, but that they were designed to mimic our screams and winces. (Block 1978: 281)

After the necessary changes have been made, Block's point applies in an obvious manner to the case of Oscar. The conclusion of the analogical argument—that Oscar must have an inner life much like our own—is undermined by the observation that Oscar was built so as to mimic our behavior.

6.5.3.2 The (Thought) Experimental Solution: Trying on Oscar

Robert Cummins has defended the bold claim that questions like the one whether Oscar is conscious can, in principle, be definitively settled by an experiment. All you need to do is to put yourself in a position in which you can experience the qualitative feel of Oscar's mental life. That is, you 'try on' Oscar and see for yourself what it is like to be him. (See Cummins (1984: 791-798))

Cummins sets up the thought experiment in the following way: we have a person, A, A's brain, B, and a computer, C. C is programmed so as to perfectly simulate the I/O of B. Both B and C receive all the sensory input from A's body. A has a switch with two positions: in the one position B controls A; in the other position C controls A. If A can tell no difference in what it is like to be him- or herself when flipping the switch back and forth, then we have a proof that the theory we used in programming C is the correct theory of the human mind.

It is obvious how we could utilize this setup to find out whether Oscar was conscious or not. You are A, B is your brain, and Oscar is C. We hook you up in the way Cummins suggested. Then you will be able to find out what it is like to be Oscar. It might be exactly like being you, it might be like something but different from being you, or it might turn out that it is like nothing at all to be Oscar.

Must we conclude that the puzzle about phenomenal consciousness is just another philosophical pseudo problem that vanishes when exposed to the harsh light of empirical inquiry? I doubt it. Though much taken with the 'trying on' approach to testing theories of the mind (see Section 3.1), I take this thought experiment to show that trying on a theory can be tricky. The problem is this. In order to make sense of the way in which the outcome of the experiment is described, we must assume that there is *one* person (or center of consciousness) who can determine whether the subjective feel stays the same throughout the switching procedure. But, given the setup of the experiment, we are not entitled

to this assumption. The right way of counting persons in this experiment yields the answer that we are dealing with *two persons that time-share one body.*

To describe this situation correctly, we must introduce another symbol, D, to stand for the debrained body of A. Now we can name the complex D&B 'Natural'; and the complex D&C 'Artificial.' Because they share (the better part of) one body Natural and Artificial are, of course, very close to each other. But they are not in a privileged position vis-à-vis the question what it is like to be the other. The problem of other minds is just as acute for Natural and Artificial as it is for any other pair of persons. The experiment conceals this fact by setting up a situation that we might want to describe thus: *One* person has or experiences *two* minds. This one person can then inform us what these minds are like, from the inside, so to speak. What misleads us into believing that the experiment actually succeeds in creating this situation is the fact that we are dealing with only one body. In normal circumstances sameness of body indicates sameness of person. But under the special circumstances of the experiment the number of bodies is not a trustworthy indicator of the number of persons. Thus, the right way to describe the situation is not: *One person* has or experiences *two minds*. What we should say is: *Two persons* share *one body.*

If Natural and Artificial really are as distinct from each other as any other pair of 'normal' persons, then they should not be able to detect such facts as that they have inverted color spectra, for example. At first sight, it may seem absurd to maintain that they could not detect such differences in the experimental situation described. Surely, A would experience the most unsettling color inversion every time the switch was flipped. And this unsettling experience would tell A that the two minds are very different with respect to experienced color qualia. Thus, the thesis that Natural and Artificial are distinct persons who cannot know whether the other is conscious appears to be refuted.

But this objection won't stand up to scrutiny. Notice that in order to formulate it, I had to switch from talking about Natural and Artificial to talking about A. This switch allowed me to smuggle in the assumption that there is really only one person who can experience two minds. If we describe the situation correctly, it becomes obvious that there exists no *one* person who experiences an unsettling color-quale inversion upon flipping the switch. Natural experiences the meadow in which D is situated as looking just the way meadows always look, and after the switch is flipped, Artificial notices nothing peculiar about the color of the grass. To both of them the grass looks the way green things normally look to them. In this sense they are both appeared to greenly. This is not to say that Natural and Artificial are having the same color qualia: in some objective sense there is an important (noncomparative) difference between them. It's just that neither Natural nor Artificial is in a position to detect this difference. While

B is doing the 'minding' for whoever is the current subject of our experiment, everything looks perfectly normal to that subject. And when control is passed to C, everything looks the way it always used to look to the subject we are now dealing with. Who is to notice the change?

Thus we get the result that you will not notice anything unusual when experimenting with Oscar, *even if Oscar should be totally unconscious.* While B (your brain) controls D (your debrained body), everything will look normal to Natural (you), and the disembodied C (Oscar) will be minding his own business. When control over D is passed to C, Artificial will carry on his unconscious existence as qualitative zombie without missing a beat. In the meantime B will continue to 'give off' qualia appropriate to the inputs it keeps receiving from D. But this phenomenology is yours, not Oscar's. Only if you mistakenly believe that you are, at this moment, somehow 'sampling Oscar's mind' could the illusion arise that this procedure established Oscar's consciousness.

I conclude that 'trying on' Oscar won't help to determine whether Oscar is conscious or not. The experimental (or 'thought-experimental') way of determining whether Oscar is conscious cannot provide an answer to our question. In the face of the genuine possibility that Oscar is thoroughly blindsighted we should continue to hold that he is unconscious.[91]

6.6 Summary

Nagel has written that "consciousness is what makes the mind-body problem really intractable." (Nagel 1974: 165) And Pollock agrees when he says:

> The single most perplexing problem for any materialist theory of the person is that of making sense of consciousness. I think it will be generally granted that we could construct an "intelligent" machine that would behave like, and in all physical respects work like a human being, but the nagging question remains, "Would it be conscious?" Would we be *creating* sentience or only *simulating* sentience? (Pollock 1989: 28)

Oscar is Pollock's answer to this question: it is impossible to create a machine that embodies the human rational architecture without, at the same time, creating a conscious machine. I have argued that the claim that *one cannot have rationality without consciousness* is false. Additionally, I have argued that we have good reasons to take the possibility that Oscar is unconscious seriously. My argument falls short of a proof that Oscar is unconscious. Much less does it show that Oscar cannot be conscious. I think of Oscar as an intelligent machine that simulates sentience. Adapting Armstrong's wonderful metaphor that pictures conscious-

ness as the cream on the cake of mentality, I can summarize my conclusion by saying that Oscar has the cake without the cream.

For the time being this brings my case against introspection to a close (the last installment has to wait until Chapter Eight). I have argued that the main reason for accepting introspectionism does not stand up to scrutiny (Chapter Five), and in this chapter I have tried to show that one of the most interesting introspectionist theories of consciousness does not hold what it promises. New and better reasons for introspectionism might be discovered; new and improved versions of introspectionism might be proposed. True. But since we cannot live from hope alone, we will be well advised to explore other possibilities in the meantime.

CHAPTER SEVEN

Relocating Qualia

To be conscious is to have qualia. In order to unpack this slogan, we have to address at least two questions: What are qualia? (The nature-problem) and What is it to have qualia? (The having-problem). The two preceding chapters addressed the second of these questions. I presented introspectionism as the prevailing theory of what it is to have qualia and argued that it should be rejected. In Chapter Six I focused on Pollock's specific version of introspectionism and attempted to show how it fails as an account of qualitative consciousness. In Chapter Five I mounted an indirect attack on introspectionism by arguing against a tempting but mistaken reason for embracing introspectionism. Since the discussion centered on the question of what it is to *have* qualia, little or nothing was said about the nature of qualia. In this chapter the focus switches to the nature-problem.

The discussion so far relied on a relatively unarticulated notion of qualia. Unfortunately, one's failing to spell out clearly what one takes qualia to be is not at all the same thing as one's failing to make assumptions about the nature of qualia. One such unspoken assumption about qualia that remained largely unquestioned throughout the many turns of the anti-introspectionist argument was the assumption that qualia are mental paint, maybe even Mental properties. Introspectionism seems to presuppose the view that qualia are, in some sense, inner—how else could *intro*spection be a means of accessing them? My critique of introspectionism did not challenge the assumption of the innerness of qualia. The assumption is palpable throughout the discussion of the belief thesis. Recall that I challenged the belief thesis on the basis of the phenomenologically established fact that there are detached qualia. I held that it is a phenomenological fact that there are qualitative mental states that are unaccompanied by corresponding qualitative beliefs. While the first appeal to phenomenology was not presented with the aim to address the nature-problem, it is quite naturally (though mistakenly) read as confirming the view that qualia are properties of peculiarly Mental items that form the natural domain of introspection. The assumption of innerness was also evident in Chapter Six. Oscar's qualia, if he has any, are of the mental

paint variety. In turning to the nature-problem, this assumption of the innerness of qualia will be the focus of the investigation.

7.1 Three (or Four) Ways of DeMenting Qualia

Qualia have often been seen as paradigms of properties that are Mental in the sense of being nonphysical. This idea is best understood as affirming that only nonphysical substances can exemplify qualia. Sense-data, as they are popularly understood, are the premier candidates to bear qualia so understood. That this association of qualia with anti-physicalism is still well and alive is borne out by a fact with which every qualophile is familiar. More often than not the admission of qualophilia is interpreted as an admission of anti-physicalism. And this is so despite the fact that there are few philosophers who defend the Mentality of qualia.

The new way of thinking about qualia opposes the traditional association of the phenomenal with the nonphysical. The aim is to free qualia from the 'guilty' association with the Mental. Hence the name of the thesis: the thesis of deMented qualia. There are various ways to deMent qualia. The most radical way of all is to deny that qualia exist. This is the thesis of the absence of qualia. A little less radical is the claim that qualia exist but that they are nothing but the ordinary perceptible properties of the external physical objects we perceive. This is the wholehearted relocation thesis (WRT). The third way to deMent qualia appears to be the most conservative. It recognizes that to deMent qualia is not necessarily to make them nonmental. According to this view, qualia are physical properties of those physical states that are our mental states. They might, for example, be neurophysiological properties. This is the halfhearted relocation thesis (HRT); I will also use the label 'colored-brain thesis' to refer to this position. So the thesis of deMented qualia can be articulated in three distinct forms: there is the absence thesis and then there are the two relocation theses. All three of these theses are versions of the thesis of deMented qualia because they deny that qualia are exemplified by Mental items. WRT and HRT are relocation theses because they construe the deMented qualia as properties of new nonMental bearers. The absence thesis deMents qualia without relocating them. In denying the existence of qualia it does make them into properties of nonMental bearers.

These three theses will occupy us in this and the following two chapters. But this may be the right place to mention a fourth possible way of deMenting qualia. It resembles the relocation thesis in acknowledging the existence of qualia. And it resembles the absence thesis in not assigning the qualia to a new nonMental bearer. In short, it is an attempt to develop a noneliminativist version of a thesis

that deMents qualia without relocating them. But the development of this thesis will have to wait until Chapter Ten where I present my own account of what qualia are and how they are had. It may be helpful to arrange these four ways of deMenting qualia in a table:

Table 1. Four ways of deMenting qualia

	noneliminativist	eliminativist
relocationist	WRT, HRT	
nonrelocationist	(Chapter Ten)	absence thesis

By relocating the traditional qualia of the sense-datum theory we change them in two important ways. Mental silver-screen qualia turn into deMented celluloid qualia. A sense-datum theorist is prone to hold that an unrelocated quale borne by a sense-datum is eo ipso had by the person to whom the sense-datum belongs. That is, sense-data exemplify silver-screen qualia. Relocated qualia are typically taken to be of the celluloid variety. They must be represented in order to be had by somebody. This is no less true of the mental paint qualia of the HRT than of the physical paint qualia of the WRT. This neat arrangement of relocated celluloid qualia on the one hand and unrelocated silver-screen qualia on the other does not show that the two distinctions are the same. The view I will defend in Chapter Ten relies on silver-screen qualia that are unrelocated. Note that 'unrelocated' does not mean the same as 'Mental.' The qualia I employ are unrelocated not because they are exemplified by Mental qualia-bearers but because they are not born by anything. Since a relocated quale is one that is exemplified by a nonMental bearer, unborne qualia do not count as relocated qualia. I also believe that the notion of an unrelocated celluloid quale harbors no contradiction. The claim I made above that an unrelocated quale borne by a sense-datum would eo ipso be had may seem to contradict the alleged independence of unrelocated and silver-screen qualia. But this is a mistake. Qualia borne by a sense-datum are automatically had not because they are unrelocated but because they are borne by a very special Mental item. The notion of a sense-datum is the notion of an object that bears the qualia one has. I see no principled objection to the idea of a Mental qualia-bearer that does not automatically present its qualia to the person to which it belongs. Having noted this complication, it must be noted that tendency to regard relocated qualia as celluloid qualia is extremely entrenched. All of the relocationists we have been and will be discussing opt for relocated celluloid qualia. We will, therefore, not go wrong in assuming that we speak of celluloid qualia whenever we speak of relocated qualia.

In the bulk of this chapter I introduce three versions of the thesis of deMented qualia, present the evidence in their favor, and draw out some of their less welcome consequences. I begin with a discussion of the absence thesis. We will see that the most famous exponent of the absence thesis is more profitably viewed as a champion of wholehearted relocation (relocation$_w$) of qualia. A brief description of the WRT is followed by a discussion of an odd consequence of the relocation$_w$ of qualia: if you push the qualia out of the head, consciousness seems to follow right along. Thus one gets what may seem like an unacceptable inflation of the sphere of consciousness. I then turn to a discussion of two arguments in support of the relocation theses. First, one can argue that qualia must be wholeheartedly relocated because that is what experience teaches. I spend considerable time with this impressive appeal to phenomenology and conclude that appeals to phenomenology cannot settle questions concerning the location of qualia. Less time is spent with the second argument which simply states that qualia must be relocated (in one way or another) because there are no Mental qualia-bearers. This point naturally leads to a consideration of the HRT, the colored-brain thesis. The obvious problem with the halfhearted relocation (relocation$_h$) of qualia, viz., that the brain is not colored in the right way, is briefly addressed at this point. The better part of the material on the two relocation theses is expository and exploratory rather than critical. The critique of the relocation theses is saved for Chapter Nine.

The chapter ends with an attempt to assess the bearing the relocation$_w$ of qualia has on the argument against introspectionism presented in the previous chapters. I try to show that the WRT has no significant consequences for the case against introspectionism.

7.2 Denying Qualia

Well aware of the counterintuitive nature of his thesis and of the danger of being accused of "extrapolating from an abnormally impoverished mental life," (Dennett 1979: 97) Dennett makes the following concession:

> at first sight my theory's domain of direct, immediate consciousness seems catastrophically underpopulated: there are no colors, images, sounds, gestalts, mental acts, feeling tones or other Proustian *objects trouvés* to delight the inner eye; only featureless—even wordless—conditional-intentions-to-say-that-*p* for us to be intimately acquainted with. (Dennett 1979: 97)

Images, sensations, impressions, raw feels, or phenomenal properties are banned from the inner theater. Propositional episodes, or judgments, understood

as "the momentary, wordless thinkings or convictions" (Dennett 1979: 95) are the stuff our mental lives are made of. As Dennett puts it:

> I am left defending the view that such judgments *exhaust* our immediate consciousness, that our individual streams of consciousness consist of nothing but such propositional episodes...There are public reports we issue, and then there are the episodes of our propositional awareness, our judgments, and then there is—so far as introspection is concerned—darkness. (Dennett 1979: 95)

This, then, is the thesis of the absence of phenomenology: our inner lives consist of nothing but judgments. There is in the inner theater no room for qualitative mental phenomena.[92] How does this thesis bear on the question Dennett can ask so evocatively: "Is that special inner light really turned on, or is there nothing but darkness inside?" (Dennett 1987a: 161) Dennett's exposition of the absence thesis might lead one to believe that he wants to convince us that we don't experience colors, sounds, tactual feels, smells, bodily sensations, etc., i.e., that there is nothing but darkness inside. But this is surely absurd.[93]

I think that Dennett's presentation of his own theory makes it sound more radical than it really is. In the paper from which I have been quoting he only gives us a hint as to where he wants to find a home for colors, sounds, etc. Discussing the experience of phosphenes—e.g., the experience of an undulating colored checkerboard that results from pressing your finger on your eyeball—Dennett writes:

> ...why should *these* phenomena seem to raise difficulties not already encountered in the normal phenomena of perception, e.g., seeing an undulating colored checkerboard in front of me? Presumably the answer is that in the latter case I can answer the question: 'So where are the square colored things?' straightforwardly and without embarrassment: 'They are about three feet in front of your nose...' but in the former case, *which is just about like* seeing an undulating checkerboard (like enough so that the same question is appropriate, presumably), I have no ready answer; there are no public, physical, external (or even internal) square colored things to point to. (Dennett 1979: 107)

Here we get at least an indication as to where Dennett wants to locate the colors, sounds, smells, etc., with which we become acquainted in veridical perception. They are somehow on or in the physical object perceived.[94] In his more recent writings Dennett has affirmed this view, arguing that colors—the only colors there are—are properties of external physical objects. These colors are, admittedly, very unruly, open ended, disjunctive properties that are virtually impossible to specify. But while this untidiness may displease the philosopher of sci-

ence, it has no tendency to show that these color properties aren't real, physical properties of real, external objects. (See Dennett (1991: Chapter 12))

This account of color suggests answers to two otherwise puzzling features of Dennett's theory. First, we don't meet colors, sounds, etc., in introspection because they are not in the mind. Second, it is not dark inside because in perception we become acquainted with the colors, sounds, etc., as they exist on the external physical objects. These outer phenomena throw their light into the inner realm of our minds; thus the inner is illuminated, albeit only by reflected light. That is to say, there is no need for a second set of colors to 'gild and stain' the mind. In the mind there are judgments, reactive attitudes, etc. The colors we experience, the only colors there are, are borne by the physical objects we perceive. If this is a correct interpretation of the thesis of the absence of phenomenology, then the absence thesis is virtually indistinguishable from the WRT, to the discussion of which I now turn.

7.3 Wholehearted Relocation

Presentations of the absence thesis sometimes create the misleading impression that it denies the rich phenomenology of perception. Proponents of the relocation theses affirm that perception acquaints us with a multitude of qualitative phenomena. But they argue that there is nothing Mental about these qualities. The wholehearted relocationist (relocationist$_w$) takes the qualia enjoyed in perception to be the qualities of the external objects we perceive. The so-called qualia are not mysterious inner qualities but simply physical properties of the perceived object. We are dealing here with a *relocation* of qualities: the qualities that were, mistakenly, taken to qualify some item in the perceiver's mind (or Mind) are now discovered to adorn the external perceived objects. The qualities are moved out of the mind and onto its objects. Once perceiving is so understood, perceptions appear to be devoid of all intrinsic phenomenal qualities.

Before turning to an evaluation of the arguments in support for the relocation theses (in Section 7.4), I want to discuss what appears to be a disturbing consequence of the switch to physical paint. It has been argued that moving qualia out of the head is tantamount to moving qualitative consciousness out of the head. While this may not amount to a *reductio* of the WRT, it does make it considerable harder to believe.

7.3.1 The Expanding Sphere of Consciousness: The Inflation Problem

Qualia were not the only and not the first items to be kicked out of the mind. Certain contents of propositional-attitude states were the first evictees. Content

externalism—enshrined in Putnam's slogan: "'meanings' just ain't in the *head!*" (Putnam 1975: 227)—has become something of a philosophical orthodoxy, notwithstanding the metaphysical problems that this view generates. A vivid way to get a grip on this issue is to ask for the location of propositional-attitude states whose content is not in the head. Many have understood externalism as suggesting the metaphysical picture "of the mind as precariously suspended between the person and these spatio-temporally removed entities—as if spatio-temporally distributed." (McGinn 1988: 21) A willingness to countenance the idea that our thoughts, beliefs, desires, etc., are not confined to our heads is often combined with an insistence that our qualitative experiences are intrinsic. Pains, tickles, phenomenal seemings, etc. are not usually subsumed by externalism. And there are good reasons for that. The most persuasive arguments for content externalism—the twin-earth arguments—seem to be based on the assumption that there is something—viz., how things seem to you—that is invariant between twin worlds. In other words, these arguments for externalism presuppose an internalism about seemings. A similar point is sometimes made using the terms 'narrow content' and 'wide content.' Many have argued that you cannot have wide content without narrow content. And some have explicitly tried to identify narrow content with phenomenology. Another consideration suggesting the combination of externalism about propositional-attitude states with internalism about phenomenal states is epistemic in nature. The epistemic fact from which we start is this: we can be and often are mistaken about what we think, believe, etc. We are rarely mistaken about our qualitative states. Our fallibility about the former states is traceable to their involving external objects of the environment. And it is a plausible conjecture that our reliability concerning our qualitative states is due to their being intrinsic—their being independent of external circumstances.

But with the advent of the WRT this distinction between propositional-attitude states and qualitative states falls. Tye has complemented Putnam's slogan with his own adapted version of it: *"phenomenology ain't in the head."* (Tye 1995a: 151) Somewhat less condensed: "the phenomenal or felt aspects of our mental lives…are not even in the head at all." (Tye 1995a: xi) And Dretske has put it into the simplest terms possible: "the mind isn't in the head." (Dretske 1995: 38) These quotations suggest a radical picture of the location of qualitative consciousness. They present externalism as, quite literally, a consciousness expanding position. The cartoonist's thought bubbles turn out to be the sober truth. The Montgolfier affords another suggestive image: the head is represented by the small gondola out of which grows what must count as a strikingly apt representation of the mind—a huge hot-air balloon. The size of your consciousness dwarfs the size of your head. How big is the sphere of consciousness? That de-

pends. Its size is largely determined by what you are looking at. If you squint at the tip of your nose your mind barely bulges beyond the barrier of your brow. But if you are more ambitious you may want to look at the Milky Way, for example, and your mind will explosively expand out of your head to measure many light years across.

This vision of the place of consciousness in the world is not new. We find it, for example, in the writings of the American New Realists. Edwin B. Holt unflinchingly faces up to this issue when he identifies a person's qualitative consciousness with the (relocated) qualia to which she reacts. Holt writes:

> Consciousness is extended in both space and time:—in space as spatial objects are extended, consciousness being actually such parts of the objects as are perceived; i.e. such parts as are consciousness; and in time as a quarter-hour, a day, or a week, is extended. (Holt 1914: 210-211)

Disregarding the dimension of time, this means that your visual consciousness is spread out over, or is identical with, the relocated qualia on the surfaces of the objects you see. Distinguish Holt's claim from the less controversial view that the manifold of physical objects with which vision acquaints us is extended in physical space. On Holt's view, consciousness itself expands out of your head like an inflated balloon until it becomes spatially coextensive with its objects.

What should one make of this peculiar consequence of the WRT? The unease one feels in the face of this strange emanation of consciousness into the world at large cannot be traced to a failure of phenomenological adequacy. On this count the thesis is in the clear. Nevertheless, one is inclined to concur Jaegwon Kim's laconic verdict: "This sounds absurd, and it is absurd." (Kim 1996: 200)[95]

7.3.1.1 Rarefaction

The most elegant way of dealing with this problem would be to deny that the relocation$_w$ of qualia gives rise to the inflation of consciousness. And it seems that there is a natural way of thinking of the relocation of qualia that does not even suggest this unwanted expansion of consciousness. One might argue that the inflation problem is an artifact of a mistaken intuitive picture of the mind.

If you picture the mind as ball of gooey substance in which are embedded its objects, then their relocation becomes a messy affair. You end up with sticky glutinous threads stretching from the mind to its externalized objects. The mind thus changes its shape from a small ball in the head to that of a giant squid whose arms reach out and grab its relocated objects. So we end up as cephalopods of sorts with mental tentacles extending out of our heads. But if we start with a different imaginative picture of the mind, we avoid this unsavory inflation of the mind. Imagine that the mind is a nice, clean box containing its objects.

Relocating these objects and turning them into physical features of external objects does not affect the shape or the size of the box. The objects are now outside and the mind is inside safely ensconced in the bulwark of the head. Inflation never happens.

The box-shaped mind has problems of its own. The most pressing question is surely this: How does the mind in here stay in contact with its objects out there? Sections 8.4 and 8.5 of the next chapter are devoted to this question. But questions remain, even if we set aside this big worry for now. Here I want to focus on the following issue. If all the objects are removed from the box then nothing remains in the mind. This leads to an unacceptable rarefaction of the mind—it makes consciousness invisible by making it completely transparent. Thus we lose sight of consciousness in the attempt to contain its inflation. I now turn to a somewhat less metaphorical discussion of this issue.

The founders of Australian materialism—U.T. Place, J.J.C. Smart, and D.M. Armstrong—are responsible for the current popularity of the transparency thesis. The substance of the thesis (if not the label) is clearly contained in Section V of Place's path breaking paper of 1956. Smart appeals to the transparency of the mental as if it were an obvious fact when he speaks of the "singular elusiveness of 'raw feels'." (Smart 1959: 61) Armstrong endorses the view that "the mental, as we actually experience it introspectively, is elusive, hard to pin down, as it were transparent or diaphanous." (Armstrong 1984a: 158) And he does not restrict the scope of the thesis to the relatively uncontroversial cases of "belief, thought, purpose and cognate mental phenomena." Armstrong makes it quite clear that he means to include the so-called qualitative states like "perceiving, imagining, the having of bodily sensations and emotions." Singling out the case of perception he says: "I will argue that *perception*, as we experience it introspectively, is entirely qualityless." (Armstrong 1984a: 169-170) Testimonials could easily be multiplied. Claims to the effect that consciousness is transparent, elusive, diaphanous, abound in the literature. The transparency of consciousness is accepted as a truism by those who think of qualia as physical paint. An appeal to experience is often taken to be sufficient to establish the physical paint scenario and the consequent transparency of the mind. I will present and critique this phenomenological argument in Sections 7.4.1 and 7.4.2.

But the transparency thesis may also be seen as arising from a certain way of thinking about the structure of consciousness. G.E. Moore's analysis of sensations is a paradigm of this way of approaching consciousness. To show this, I will briefly present Moore's position and then state my reasons for doubting the purported truism of the transparency of consciousness.

What makes the transparency thesis seem so obvious? I think it is a line of thought about consciousness clearly set out in G.E. Moore's influential paper

"The Refutation of Idealism." (Moore 1903) Moore distinguishes two parts in every sensation: (i) awareness (or consciousness), which is the same in all sensations; and (ii) the object of awareness, which is different in different sensations. Take, for example, a sensation of blue and a sensation of green. According to Moore, these two sensations are identical in the element of awareness, but they differ in their objects. The object of the one is blue, that of the other green. Now how do I 'get at' the blue—the blue that is the object of my sensation and, according to Moore, not in my mind? How does my awareness become an awareness *of* the blue? What, to put the question into my language, lets me *have* the blue in such a way that, as a result, it becomes like *that* to be me? Moore does appear to answer this question. He speaks of a special kind of relation—he calls it 'knowing' but he might as well have called it 'direct awareness'—the obtaining of which makes my consciousness be a consciousness *of* blue. By thus *knowing* the blue I come to have it in a way that results in its being like *that* for me. Here is how Moore introduces this relation: "And this awareness is not merely…itself something distinct and unique, utterly different from blue: it also has a perfectly distinct and unique relation to blue…This relation is just that which we mean in every case by 'knowing.'" (Moore 1903: 24-25) Discussion of the question how satisfactory Moore's solution of the having-problem is will have to wait until Section 8.5. Here I want to pursue a different issue. Moore's account suggests the 'elegant solution' to the inflation problem of consciousness. But it also lets us see the price of pursuing this route: it threatens to make consciousness so rare that we cannot grasp it.

After splitting every sensation into two components Moore identifies consciousness with the element that is common to all. This subtraction from consciousness of all external admixtures creates the impression that the purified remainder withdraws into its point of origin thus reversing the inflationary expansion of consciousness. While welcoming this contraction, one cannot help but wonder about the nature of this rarefied remnant of consciousness. Moore is well aware of this worry. He speaks repeatedly of the difficulty involved in getting consciousness as such into focus. Introspectively, he says, consciousness "is extremely difficult to fix." He thinks that many people "fail to distinguish it at all" and, metaphorically speaking, he attributes this to the fact that consciousness is "transparent—we look through it and see nothing but the [object of consciousness]." (Moore 1903: 20) Consciousness "is as if it were diaphanous." (Moore 1903: 25)

In this way consciousness has been turned into something hidden by a process of rarefaction that has made it invisible. It now presents itself as a theoretical posit—we postulate it based on theoretical considerations—rather than as a pre-analytic datum. You will recall that I started this investigation into qualitative

consciousness by emphasizing the blatancy of qualitative consciousness. We are qualitatively conscious at every moment of our waking and dreaming lives. And at every moment of (at least our waking) lives we can easily become aware of this fact. So Moore's claim seems quite wrongheaded to me. I can do no better than to agree with William Robinson who has put the point most forcefully:

> Moore was dead wrong when he said that consciousness is "diaphanous"...Every time we have a pain or an afterimage, or have our attention arrested by the color, sound, smell, feel, or taste of something, we know one or more of the kinds of consciousnesses. We cannot complain that we thereby do not know what consciousness is. (Robinson 1994: 13)

Here is how I would analyze what (mis)led Moore and his many followers to embrace the idea that consciousness is elusive or transparent. Consciousness is the having of qualia. That is, for there to be any sort of qualitative consciousness at all, there needs to be a quale that is being had by someone. The quale alone and the having-relation alone are insufficient to ignite the light of consciousness. Moore insists on isolating the having-relation alone and calling it consciousness. This is a mistake. But I have to agree with him that these pure havings are quite intangible.[96]

The upshot for our discussion of the inflation problem is this. If the relocationist follows Moore and identifies consciousness with the having alone, then she will not be forced to extend consciousness outwards to the external quale-bearing object. But this way of containing the sphere of consciousness within the subject comes at a price. Simply put, we now no longer know what it is that has thus been reigned in. It is a diaphanous, elusive something—if it is anything at all—that is a far cry from the full blooded spectacle of qualitative consciousness that engulfs us at all times. I conclude that a Moorean account of consciousness is an unpromising way to stop consciousness from spreading throughout a world filled with the physical paint of relocated qualia. The Moorean way of interpreting the metaphor of the box-shaped mind shows that this natural way of blocking the unwanted inflation of consciousness is unsuccessful. The rarefied sort of consciousness that this maneuver manages to confine to the box of the mind is not the qualitative consciousness we are after. Thus, the problem of the inflation of consciousness still remains to be addressed by the relocationist$_w$.

7.4 Support for the Relocation Theses

Support for the relocation theses comes in two parts, one positive the other negative. On the one hand, their proponent needs to provide positive reasons for

accepting the relocation theses. On the other hand, we need to be given reasons that allow us to discard what seem to be crushing objections against the relocation theses.

Positive support for the relocation theses flows from two sources. First, the champions of the WRT hold that the testimony of experience shows their thesis to be true. Second, it is alleged that the friend of qualia cannot provide a home for phenomenal qualities in the Mind. Relocating these properties wholeheartedly or halfheartedly is, therefore, the only way to save them from elimination. The discussion of these arguments will lead straight into some of the gravest difficulties the relocation theses have to face.

The two most prominent objections against the relocation theses are the argument from illusion and the Democritean argument from atomism. As long as these objections stand, the relocation theses fall. The argument from illusion focuses on cases of nonveridical experience where there is no external physical object that can plausibly be said to bear the qualia one experiences. The argument from atomism is directed against the claim that the perceived physical object is a promising new home for wholeheartedly relocated phenomenal qualities. This argument is not restricted to cases of nonveridical experience and cuts equally against the WRT and the HRT. Based on the best extant accounts of the nature of physical objects, this argument purports to show that *no* physical object—neither the tomato you see nor the neural state that is your seeing of it—is a fitting bearer of the relocated qualia.

To provide the negative support for the relocation theses, one has to overcome these two objections. No matter how good the positive arguments for the two theses may be—as long as these two objections have not been silenced, neither WRT nor HRT can get off the ground. In this chapter I will confine myself to a discussion of the positive reasons for accepting the relocation theses. The discussion of the two objections against them will be postponed until Chapter Nine.

7.4.1 The Second Appeal to Phenomenology

While the first appeal to phenomenology (Section 5.9.2.1) did not explicitly assert the inner nature of detached qualia, it now appears clear that their 'innerness' was a tacit assumption behind this appeal. The second appeal to phenomenology—one of the pillars of relocationism$_w$—brings this tacit assumption into relief and provides reasons to reject it. In this new light the probative force of the first appeal to phenomenology may therefore stand in need of reevaluation.

The most impressive support for the WRT comes from a simple but powerful appeal to experience. Every time one opens one's eyes the truth that qualia are

outside presents itself with all but irresistible force. So the relocationists find themselves in the enviable position of defending a thesis whose obvious truth is borne out in every casual glance. To see, one might say, is to see that the WRT is true. Here is how Armstrong states the point:

> I look out of my window and see the green leaves of a vine. Color, once we give it our attention, is a very striking and arresting phenomenon. But it does not seem to be a mental phenomenon. If we consult perception, then its verdict is clear. The green color is a property of the vine leaves. Nor is it a property which involves any relations that the leaves have to any other object, whether objects in the field of view or the perceiver. It is an intrinsic property of the leaves. In what follows I will hold fast to this perceptual deliverance. (Armstrong 1984: 170)

Gilbert Harman argues in the same vein as Armstrong. To claim that in visual perception we are aware of 'mental paint' rather than of color properties of the objects of perception is, as he puts it, "counter to ordinary visual experience." (Harman 1990: 39) He elaborates the point as follows:

> When Eloise sees a tree before her, the colors she experiences are all experienced as features of the tree and its surroundings. None of them are experienced as intrinsic features of her experience. Nor does she experience any features of anything as intrinsic features of her experience. And that is true of you too. There is nothing special about Eloise's visual experience. When you see a tree, you do not experience any features as intrinsic features of your experience. Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree...(Harman 1990: 39)

Harman's way of stating the case for relocation$_w$ also has the virtue of explicitly emphasizing the companion thesis of the physical-paint scenario. That is the claim of the transparency of the mental (see Section 7.3.1.1), i.e., the claim that there is no mental paint. *Prima facie* the appeal to the experiential validation of the WRT does seem highly plausible. But further consideration of the issue will, I hope, erode this assurance.

7.4.2 The Inconclusiveness of the Appeal to Phenomenology

A remark attributed to Wittgenstein can serve as point of departure for the critique of the relocationist's appeal to phenomenology:

> Meeting a friend in a corridor, Wittgenstein said: "Tell me, why do people always say it was *natural* for man to assume that the sun went round the earth rather than that the earth was rotating?" His friend said, "Well, obviously,

> because it just *looks* as if the sun is going round the earth." To which the philosopher replied, "Well, what would it have looked like if it had looked as if the earth was rotating?" (Stoppard 1972: 75; see also Anscombe (1959: 151))

As is not unusual with Wittgensteinian remarks, its precise meaning is not at all clear. I favor the following interpretation. The rhetorical question with which Wittgenstein ends the conversation makes us see that a heliocentric world would look no different to us than the geocentric world does. Thus, the interlocutor's explanation of why people unhesitatingly inferred the geocentric thesis from their visual data is defeated: the visual data do not favor one hypothesis over the other.

The use to which I want to put this Wittgensteinian remark should be obvious. Making the necessary changes, we get the following dialogue between an unspecified philosopher and his friend:

> Meeting a friend in a corridor, the philosopher said: "Tell me, why do people always say it was *natural* for man to assume that colors were features of the perceived objects rather than features of their visual experience?" His friend said, "Well, obviously, because it just looks as if the colors are features of the perceived objects." To which the philosopher replied, "Well, what would it have looked like if it had looked as if the colors were features of the visual experience?"

The point of the philosopher's last question is that the two possibilities are experientially indistinguishable. Mental paint looks no different than physical paint. This observation defeats our philosopher's friend's explanation of why it is supposedly natural for people to assume that colors are features of perceived objects.

Let us elaborate the analogy between Wittgenstein's astronomical example and our example concerning the experienced location of qualia. On the one hand we have two astronomical hypotheses and a set of perceptual data that strongly seem to favor one of the two hypotheses over the other. But once we realize that the observations predicted by the competing hypotheses are indistinguishable from the data we currently possess, the inference from our data to the one hypothesis is defeated. On the other hand we have two metaphysical hypotheses and a set of experiential data that strongly seem to favor one of the two hypotheses over the other. But once we realize that the data the competing hypotheses would lead us to expect are indistinguishable from the data we currently possess, the inference from our data to the one hypothesis is defeated. Here is how the elements of the analogy line up:

The astronomical example	The metaphysical example
• The geocentric world view	• The physical-paint account of secondary qualities
• The heliocentric world view	• The mental-paint account of secondary qualities
• The heliocentric and the geocentric world look the same to us.	• The mental-paint and physical-paint worlds look the same to us
• The discovery that the visual data afforded us in a heliocentric world are indistinguishable from the visual data afforded us in a geocentric world defeats the inference from our current visual data to the conclusion that ours is a geocentric world.	• The discovery that the experiential data afforded us in a mental-paint world are indistinguishable from the experiential data afforded us in a physical-paint world defeats the inference from our current experiential data to the conclusion that ours is a physical-paint world.

Wittgenstein suggests that the 'obvious' explanation of humankind's tendency to interpret its astronomical observations as supporting the geocentric world view is mistaken. But he does not tell us what the correct explanation is. Without going into any detail, I suggest that the correct explanation of this bias lies in the fact that we are small, earth-bound creatures: to us as small, earth-bound creatures it will always look as though our planet were stationary and the sun were revolving around it. What is the correct explanation of our bias to interpret our experiential data as supporting the physical-paint account of colors? The appeal to the 'obvious' explanation is no more successful in the metaphysical example than in the astronomical example. One may be tempted to suggest something along the following lines: our innate tendency towards naive realism favors the physical-paint interpretation of our visual experience. Thus we can record another element of the analogy:

• Our earthbound existence biases us towards interpreting our visual data (that are really neutral as between the two alternative astronomical hypotheses) in a way that is consistent only with the geocentric world.	• Our naively realistic metaphysics biases us towards interpreting our experiential data (that are really neutral as between the two alternative metaphysical hypotheses) in a way that is consistent only with the physical-paint world.

I think that the analogy becomes a little strained when it is called upon to explain why the majority of people interpret their experience in a manner that accords with the WRT. Up to this last step each element of the analogy suggested its counterpart naturally. But this naturalness breaks down at the last step considered. The correct explanation of our bias towards geocentrism is silent when

called upon to suggest an explanation of our bias towards the physical-paint hypothesis. Aside from this unnaturalness, the proposed explanation of our metaphysical bias is weak at best. The appeal to an innate realism seems almost empty: it is, after all, merely a puffy way of restating the fact that we tend to favor the physical-paint interpretation of the appearances.

Though regrettable in itself, the analogy's eventual breakdown need not trouble us. The analogy stood up long enough to allow us to make our point against the relocationist. We called upon the analogy to undermine our opponent's claim that the WRT is simply a 'perceptual deliverance' and that it would be 'counter to ordinary visual experience' to deny it. The discussion of the analogy has shown that there is here no simple lesson to be learned from experience. What we experience visually fits both stories equally well. Visual experience cannot be the arbiter between the mental-paint story and the physical-paint story.

This result may lead the relocationist to reevaluate his initial acceptance of the analogy. Most likely it will be the third step which will appear problematic in this new light. After all, what reason have we been given to believe that the mental-paint story and the physical-paint story are experientially equivalent? Is it not plausible to assume that our experiences should be completely different in a world in which phenomenal qualities are properly classified as mental paint?

I think that those who are attracted to the WRT have a tendency to believe that if the mental-paint story were correct we should experience something like a flat image in our heads—a picture painted onto the famous 'veil of perception,' as it were.[97] But surely we have no reason to accept this view. Consideration of the following possibility may help to understand why we would experience qualia 'out there' rather than 'in here,' even if the mental-paint story were correct. Imagine a pair of special goggles. Instead of lenses they have little screens onto which is projected an image of what is in fact in front of you. If these goggles were crafted very well, you would not see two images projected onto the two little flat screens in front of your eyes. Instead, you would see a full-featured three-dimensional world in front of you.[98] You would 'look right through' the images in your goggles and be aware only of the objects represented by these images. After a while it actually might become very difficult to redirect your attention to the images themselves; they might have become completely transparent to you. Try as you will, you might not be able to focus on the intrinsic features of the images. But in this case we would surely not be tempted to infer that, therefore, there are no images. We understand how the goggles work, and we know that it is only in virtue of the images having the intrinsic qualitative features they have, that we are able to perceive anything. So our difficulty of focusing on the intrinsic features of the images does not mean that there are no images, nor would it mean that they have no qualitative intrinsic features.

An analogous account may be true of our ordinary visual experience. The fact that we see colored physical objects "a great way off" (as Hylas put it) is just what you would expect on the mental-paint story. That we cannot, at will, make our phenomenology 'snap back' into the mental picture painted with mental paint no more refutes the mental-paint account than does our inability to see the images in our goggles refute the claim that we are wearing goggles.

Paying attention to the goggle example should allay the worries of those who doubt the claim that the two alternative accounts of secondary qualities are experientially indistinguishable. Considered in their entirety, the reflections to which Wittgenstein's remark gave rise do not prove the claim that our visual experience has intrinsic phenomenal properties. But they do serve to undermine the self-assured appeal to ordinary visual experience of which the proponents of the WRT make so much. The probative force of the second appeal to phenomenology has suffered a severe setback.

It seems, then, that this first engagement with the relocationist has not helped to settle the issue of the location of qualia. Appeals to phenomenology cannot decide the issue. In Chapter Nine I will try to show that there are other reasons for believing that the WRT and the HRT are false. But at this point of the debate both, the mental-paint as well as the physical-paint construal of qualia, are live options.

7.4.3 No Mental Home for Qualia

The second positive reason for the relocation of qualia supports both versions of the relocation thesis. It is the claim that there is no room for phenomenal qualities in the Mind. Hence, on pain of elimination, a new home must be found for these properties. And it is argued that the correct locus of the qualitative properties is a nonMental object. The relocationists$_w$ opt for the external physical object of perception. The halfhearted relocationist (relocationists$_h$) choose the neural states involved in the perception of the external object.

What motivates the extrusion of the sensible qualities from the Mind? How is the metaphorical claim that the Mind has no room for phenomenal qualities to be cashed out? The relocationist's answer is best approached by asking the following question: What has it meant, traditionally, to say that phenomenal qualities are in the Mind? Many things have been meant, of course. But one popular interpretation had it that the phenomenal qualities qualify Mental items like, for example, ideas, sense-data, sensa, or sensibilia. Now the relocationist's position can be formulated simply. Qualia are not in the Mind because the Mind does not contain any of these purported Mental qualia-bearers. As the traditional occupants of this role have fallen into disrepute, it has become necessary to move the

qualia out of the Mind and into the nonMental world. And once the qualia are relegated to the nonMental world, the external physical object to which the mind is perceptually related appears to be the most likely bearer of the phenomenal properties.

How should one reply to this challenge? The most straightforward reply would consist in a defense of sense-data. Though I do not share the widespread conviction that hypothesis of sense-data has been definitively discredited, I agree with the relocationists that sense-data do not bear our qualia. But from the fact that sense-data are off limits it does not follow that relocating qualia onto external physical objects is the only way to go. In Chapter Nine I will argue at length that the physical-paint story is up against formidable problems. The relocationist$_w$ may have good reasons for assuming that the Mind is not a good place for qualia. But we will see that there are also good reasons for assuming that the mind's intentional object is not a good place for qualia either. So it appears that we get stuck with the vexing result that the Mind and the nonMental external world are equally inhospitable to qualia. In the light of this paradoxical conclusion we may have to explore somewhat more 'heroic' attempts to find a place for qualia in the world. It is in the nature of positions located in the farther reaches of logical space to offend our robust, common sense metaphysics. Faced with this tension between philosophical theorizing and common sense it would be unwise to simply say: so much the worse for commonsense. But it would be equally unwise to let common sense set the limits on philosophical imagination. If philosophical speculation is cut loose from common sense, it is apt to lead to lunacy. If common sense sets the limits on admissible philosophical speculation, philosophy suffocates. In exploring possibilities we have to try to steer between these extremes. The navigation is risky, especially since there is considerable disagreement about where these extremes start. My own proposal is presented in Chapter Ten. That it does not suffer from the suffocating stranglehold of common sense is beyond reasonable doubt. Whether everybody will want to issue this proposal a clean bill of mental health will remain to be seen. But for now I want to discuss a variant of the relocation thesis that is commonly thought to be a paradigm of sanity. I hope to show that this judgment may be somewhat too generous.

7.5 Halfhearted Relocation: The Colored-Brain Thesis

We have seen that the second appeal to phenomenology does not achieve its goal—perception does not tell us where qualia are. And if perception were to tell us that qualia are externally located it would be lying. This I hope to show in

Chapter Nine, where I argue that external physical objects are unfit to bear qualia. So qualia are not external. Nor do qualia have a Mental home. For we have agreed to grant the relocationist's assumption that the purported phenomenal objects—whose job it is to instantiate qualia—do not exist. So qualia are not exemplified outside, nor are they exemplified inside. In short, qualia aren't exemplified at all. Whether this amounts to qualia eliminativism will depend on one's account of properties. But no matter how we answer this metaphysical question, surely we can agree that unexemplified qualia are of no use in explaining consciousness. It would seem, then, that the project of explaining consciousness as the having of (exemplified) qualia has ground to a halt.

The way out of this dilemma may seem obvious: reject the false alternative upon which this purported problem is based. It is not true that sense-data and external physical objects exhaust the class of possible qualia-bearing objects. The most obvious candidate has been left out: internal, physical objects—neurophysiological states or processes—are the natural locus of phenomenal properties.

But as soon as this proposal is stated a question arises: What has become of the relocation idea? To place qualia into the brain is not to relocate them, surely. The stance one takes on this issue will depend on one's understanding of the idea of relocation. The matter can be traced back to a certain ambiguity of the terms 'inside' and 'outside' as they are used in this discussion.[99] One who denies that qualia are in the mind might be doing (at least) one of two things. He might be denying that qualia are borne by Mental (i.e., nonphysical) phenomenal objects, e.g., sense-data. Or—long accustomed to not taking sense-data seriously and much impressed by the perceptually revealed transparency of experience—he might be denying that qualia are exemplified by the brain (or anything else in the head). It is this latter denial that relocationists typically want to make. Thus the usual construal of relocation as the eviction of qualia from the head and their resettlement in the perceiver's environment. But choosing the former reading allows us to reconceive the relocation move. On that reading, the challenge that there is no room for phenomenal qualities in the mind can be met by moving the qualia out of the Mind (off the sense-data, so to speak) and into the brain. We end up with qualia that are deMented but mental. They are deMented because they are not borne by nonphysical sense-data. They are mental because they are borne by such physical states as are also mental states. Qualia thus become properties of the brain. Hence the label 'the colored-brain thesis.'

It seems, then, that the relocation of qualia can mean different things to different people. What you take the destination of migrating qualia to be will depend on your view of their original residence. If the journey of qualia starts from the brain, the brain is not a possible terminus of relocation. But if the journey starts

from the sense-datum, relocation into the brain becomes a real possibility. In counting both of these doctrines as versions of the relocation thesis I do not mean to minimize the disagreements between them. The colored-brain thesis flies into the face of the spirit of the WRT. The idea of the WRT was to get the phenomenal properties out of the mind and onto the mind's object, where this object was understood to be the external and mind-independent. By moving the phenomenal properties out of the Mind and onto the (physical) brain states, we have, at best, reached a halfway house. Moreover, to one who supports her view by appealing to phenomenology in the style of the proponents of the WRT, the colored-brain thesis must seem wrongheaded. For, just as surely as it does not look to her as if her Mind were green, it does not look to her as if her brain were green when she perceives the green leaves of a vine. These differences notwithstanding, I hope to have presented a rationale for wanting to see the colored-brain thesis as at least a halfhearted version of the relocation idea. In this spirit, I use the labels 'wholehearted relocation' and 'halfhearted relocation' to stand for the two versions of the relocation move. The relocationist$_w$ takes qualia to be properties of external, physical objects in the perceiver's environment. For the relocationist$_h$, qualia are properties of the perceiver's brain. The disagreement concerns the question whether qualia are best construed as physical or mental paint.

7.5.1 Mental Representations Old and New

The colored-brain thesis may seem to be related to the old sense-datum theory. Both theories employ mental qualia-bearing representations as their central construct. But the analogy ends here. The fear that the colored-brain thesis is just a refurbished sense-datum theory is as groundless as the hope that it might share some of the older theory's more attractive features.

Sense-data, as commonly conceived, were Mental items and the qualia they bore were of the silver-screen variety. There is considerable debate among the proponents of the HRT about the nature of mental representations and the qualia they bear. The representations might, for example, be construed as functional, computational, syntactical, or neurophysiological items. And these differences would have to be reflected in the nature of the qualia they bear. But no matter how this dispute is settled, it is clear that the resulting representations will be nonMental items and the qualia they bear will be of the celluloid sort. So the similarity between the HRT and sense-datum theory is quite superficial.

It is true that not all of the traditional sense-data philosophers took sense-data to be mental (see Section 2.2.2.4). G.E. Moore, for example, spend many years agonizing about the question whether visual sense-data are identical with the sur-

faces of perceived, external objects. Viewed from this angle the WRT looks much the Moorean version of the sense-datum theory. And before Russell gave up his belief in sense-data he, too, asserted repeatedly that sense-data were not mental but part of the physical world: "I hold strongly that the sense-datum is *not* mental—indeed my whole philosophy of physics rests upon the view that the sense-datum is purely physical." (Russell 1915: 391-392) Since Russell tended to place sense-data inside the body his view is reminiscent of the HRT. So there may be a much closer relationship between this version of the sense-datum theory and the relocation theses. But that need not embarrass the colored-brain theorists. For the deMented sense-data of Russell and Moore should not evoke the metaphysical horror of their more traditional ancestors.

In distancing himself from the traditional sense-datum theory the colored-brain theorist gains metaphysical respectability. But there is also a downside to this flight from nonphysical sense-data with their silver-screen qualia. The sorts of representations and the sorts of qualia to which the naturalistically inclined relocationist can help himself make it much more difficult to explain why having and using these qualia-bearing representations should make their user qualitatively conscious. In short, the switch from the sense-datum theory to the colored-brain thesis will greatly increase the difficulty of providing a solution to the having-problem. I come back to this question in Sections 8.2 and 8.3.

7.5.2 Qualia—"Nowhere to Be Found in the Head"

The discussion of sense-data shows that the colored-brain thesis is not just a warmed up version of the sense-datum theory. Many will want to count this as a big point in its favor. But few will think that this result suffices to establish the colored-brain thesis as true. What, then, are we to make of the colored-brain thesis? As I have already noted, most advocates of the WRT will want to reject it. But once the second appeal to phenomenology has been neutralized, relocation$_h$ begins to look like a genuine alternative. The thesis should therefore be judged on the basis of its capacity to deal with the nature-problem and the having-problem.

For the relocationist$_h$ the nature-problem becomes the question: Are qualia the sorts of properties the brain can instantiate? A serious discussion of this question will have to wait until Chapter Nine. But there is a simpler question that we can ask and answer right away: Does one's brain instantiate the qualia one has? The answer is no. It takes no great neurophysiological sophistication to discover that the brain is neither phenomenally red, green, or yellow, nor does it stink or taste of chocolate. In short, your brain does not exemplify the qualia you have.[100] This observation—the colored-brain problem from Section 2.2.2.8—is so obvi-

ous that its importance is easily underestimated. That is not to say that the point has been lost on the relocationists$_h$. They typically reply by drawing a principled distinction between the colors that adorn external physical objects and mental paint. This distinction can take different forms. One might hold that one must distinguish between the property red as it is instantiated by external physical objects and as it is instantiated by mental representations. Or one might hold that there are two sets of correlated properties—the red that the ripe tomato instantiates and the distinct but correlated red that the percept of the tomato instantiates. This strategy elegantly disposes of the obvious problem faced by the colored-brain thesis. To expect that the brain of one who perceives a ripe tomato is red in the same sense in which the tomato is red is to confuse physical color and mental paint. The brain instantiates a red quale. But a brain state's instantiating a red quale does not manifest itself by its turning physically red. Only if it did that would the brain state look to the observing neurophysiologist like a ripe tomato looks to her patient.

This way of dealing with the colored-brain problem raises questions of its own. Let us assume that the relevant property instantiated by my percept of the ripe tomato is some complex neurophysiological property. Instantiating this property lets me experience a homogeneous red expanse. The existence of this red spot in my visual field raises two questions. With the nature-problem in mind one will ask: How can this homogeneous red quale of my experience be identical with the complex neurophysiological property of my percept? With the having-problem in mind one will ask: How must I be related to the complex neurophysiological property of my percept so that, as a result, it becomes like that for me to see the ripe tomato?

These questions become more difficult as the distance between our experientially mediated grasp of qualia and our theoretically mediated grasp of qualia grows. The strategy of drawing a principled distinction between physical colors and mental paint threatens to enlarge this gap. The WRT stays close to the spirit of naive realism. The color qualia we experience simply are the colors of the physical objects we see. There is no gap here because there are no two sets of colors, no two forms of inherence, no two forms of predication, etc. This simplicity is lost in the HRT. By driving a wedge between physical color and mental paint, the theory makes it harder to grasp the notion of qualia it uses. This wedge makes it impossible to think of relocation$_h$ as being just like relocation$_w$ the only difference being that the red the latter ascribes to the tomato the former ascribes to the neural percept of the tomato. The mental paint of the HRT is not simply the internalized physical paint of the relocationist$_w$. In place of the one red (the one way of being red) we now have two. And both are less well understood than the one property they replace. As the mental paint of the HRT becomes more

elusive, the distance between it and the qualia as we experience them grows. And so the nature-problem and the having-problem threaten to become harder to answer in this framework.

7.5.3 Exemplifying Qualia and Having Qualia

This is not the place for a full discussion of the nature-problem and the having-problem as it poses itself in the context of the colored-brain thesis. But the colored-brain thesis invites consideration of the difference between the exemplification of qualia and the having of qualia. For a quale to be had (in a way that results in its being like something to be the subject who has it) it is not sufficient that it be exemplified. Reflection on wholeheartedly relocated qualia makes this clear. As long as no sentient creature sees the ripe tomato, the phenomenal red that graces its surface is not had by anybody. Surely, it isn't like anything to be a tomato that exemplifies a red quale. Nor can we assume that the nonsentient objects in the tomato's environment have the tomato's phenomenal redness in such a way that it thereby becomes something to be them. For this red quale to be had it must enter into a special relation with a potentially conscious subject, a relation such that being so related to the red quale makes it be like something of a quite characteristic sort to be the subject that is so related to the quale.

The distinction between the exemplification of qualia and the having of qualia may be patently obvious when we consider qualia that have been extruded onto external, physical objects like, e.g., tomatoes. But the point is no less valid if the relocationist stops short of pushing the qualia all the way out onto the external physical objects and locates them, instead, in some 'internal' physical objects—the retina, the optic nerve, or the brain. Because the colored-brain thesis pushes the qualia further in and thereby moves them closer to the experiencing subject, one's grasp of the difference between the exemplification of a quale and the having of a quale may become blurred. For is it not the case that the maneuver of placing the qualia into living brains guarantees that they are always at a place where there is an experiencing subject? So what is to stop one from saying that qualia that are exemplified by a living brain are *eo ipso* had by the subject whose brain exemplifies them?

This elegant transition from exemplification of qualia to the having of qualia is entirely without foundation. Even if the colored-brain thesis were true (and your brain instantiated all the qualia you had) this still would not explain how you could have those qualia. This point lies buried in an apparently naive observation that authors as diverse as Dennett and Nagel have made. The instantiation of color qualia in our brains would do us no good since, as Dennett has so perspicaciously pointed out, "it's pitch black inside your skull and, besides, you ha-

ven't any eyes in there to see colors with." (Dennett 1991: 28) So you could not see these qualia even if they were there.[101] Nagel has made the same point, if in a somewhat more unsavory manner. Commenting on the location of the gustatory qualia you have while eating chocolate he writes:

> Suppose a scientist were crazy enough to try to observe your experience of tasting chocolate by *licking* your brain while you ate a chocolate bar. First of all, your brain probably wouldn't taste like chocolate to him at all. But even if it did, he wouldn't have succeeded in getting into your mind and observing *your* experience of tasting chocolate. He would have discovered, oddly enough, that when you taste chocolate, your brain changes so that it tastes like chocolate to other people. (Nagel 1987: 30)

These examples may seem frivolous, but the point is serious. There is no reason whatsoever to believe that you are going to have (in the technical sense) the qualia that your brain exemplifies. The distinction between the exemplification of qualia and the having of qualia stands. The move from the WRT to the colored-brain thesis does not blur this distinction in the least.

7.5.4 The Prospects of Halfhearted Relocation

The colored-brain thesis is at odds with the spirit of relocationism. It adds nothing to our understanding of the nature-problem. And the hope that it offers a simple solution of the having-problem is rooted in confusion.[102] In a word, relocation$_h$ has not much to recommend itself. It suffers from many the problems that afflict relocation$_w$ while lacking its (*prima facie*) phenomenological plausibility.

Having said this, I want to end the treatment of the HRT on a more sympathetic note. Much fun has been made of, and much abuse has been heaped upon, the colored-brain thesis. My brief discussion of it is not entirely blameless on this account. On behalf of the thesis I would say that it is not at all obvious to me that the more sophisticated versions of this thesis have been refuted. In fact, it would not be too misleading to view the thesis I defend in Chapter Ten as a sophisticated version of the colored-brain thesis.

7.6 The Price of Wholehearted Relocation

The basic idea of the relocation of qualia as well as some of the most powerful considerations in its favor are now before us. We distinguished the WRT and the HRT and I provided some critical commentary on each. The relocation theses are primarily an attempt to come to terms with the nature-problem. The problems the

theses face on this count will be addressed in Chapter Nine. Notwithstanding the fact that the having-problem is not at the focus of the relocation theses, relocated qualia still need to be had for there to be qualitative consciousness. The problems surrounding the having of relocated qualia are taken up in Chapter Eight. At this point I want to pause the discussion of the relocation theses in order to address a number of pressing issues that the relocation$_w$ of qualia seems to raise for my own project.

What I want to convince you of eventually is that no one has provided a satisfactory account of how relocated qualia are had. No one has managed to explain why relating yourself to your relocated qualia in a particular way makes it be like something to be you having your qualia in this way. But it may seem that the most pressing issue at this juncture is not a critique of other people's work. I should instead be concerned about the consequences that the relocation$_w$ of qualia has for my own position and for the anti-introspectionist argument presented in the previous chapters. I believe that the relocation move leaves the dialectical situation unchanged. The impression that the externalization of qualia has a momentous impact on my work is traceable to a misunderstanding of the aim of my project and the scope of the anti-introspectionist arguments. The clarification of these issues will bring into sharper focus the nature of the claims I am making. Specifically, I want to look at three of the purported consequences that I have to confront. First, it may seem that the relocation$_w$ of qualia provides a short and decisive refutation of the belief thesis. Second, it may seem that the relocation$_w$ of qualia undercuts my argument against the belief thesis. Third, it may seem that the relocation$_w$ of qualia pulls the rug from under the project of explaining consciousness in terms of qualia. It remains to be shown that the relocation of qualia has none of these intuitively plausible consequences.

7.6.1 Does Wholehearted Relocation support the Belief Thesis?

In Chapter Five I mounted an indirect attack on introspectionism by attacking its main source, viz., the belief thesis. The case turned on the issue of detached qualia. According to the belief thesis, there cannot be detached qualia. I argued that the belief thesis is false because there are detached qualia. It took considerable work to establish this existence claim. The present suggestion is that the existence of detached qualia follows trivially from the relocation$_w$ of qualia, thus providing an elegant refutation of the belief thesis. I wish that this were so, for it would considerably simplify my case against introspectionism. Regrettably this short route to victory is not viable.

Here is the simple piece of spurious reasoning: Isn't it *obvious* that the wholeheartedly relocated qualia that grace the surfaces of nonmental, external,

physical objects are detached qualia? But the existence of detached qualia falsifies the belief thesis. So, if the WRT is true, the belief thesis must be false.

Unfortunately this elegant refutation of the belief thesis rests on a confusion. In stating that qualia are exemplified by nonmental, external physical objects the WRT makes a claim about the nature of qualia. Thus the WRT addresses the nature-problem. But it is silent on the having-problem, for it says nothing about how such externalized qualia are or can be had. The thesis of detached qualia, on the other hand, is a thesis that addresses the having-problem. It holds that having qualia is not a matter of entertaining qualitative beliefs. But it is silent on the nature-problem, for it says nothing about the nature of the qualia which, according to it, are had in a manner that does not rely on accompanying qualitative beliefs. It is, therefore, a mistake to argue that the WRT supports the thesis of detached qualia. For it is an open question whether wholeheartedly relocated qualia are detached, i.e., whether having them is a matter of entertaining the relevant qualitative beliefs. The WRT does not speak to this question.

The confusion probably arises because it is natural to think of relocated qualia as being *detached*—detached, that is, from the minds of the subjects who have them. But in this context the term 'detached' does not have the technical meaning that it was assigned by the thesis of detached qualia. Ordinarily we tend to take qualia to be mental properties, i.e., properties that qualify mental events or particulars like, e.g., sense-data. Thus wholeheartedly relocated qualia may appear oddly detached from their 'natural' bearers. But all this pertains only to the question of the locus of qualia exemplification, without so much as touching on the issue of what it is to have a quale no matter where it happens to be instantiated. The tempting inference from the truth of the WRT to the falsity of the belief thesis is fallacious.

7.6.2 Does Wholehearted Relocation undermine the Belief Thesis?

The discussion of the previous section shows that moving qualia out of the head does not falsify the belief thesis—the thesis that one cannot have a quale without having a qualitative belief about it. Could it be that the relocation$_w$ of qualia can serve to strengthen the belief thesis by neutralizing one of its defeaters?

My attempt to falsify the belief thesis (and, by implication, introspectionism) rested primarily on an appeal to the phenomenologically obvious fact that there are detached qualia (see Section 5.9.2.1). The relocation$_w$ of qualia appears to defeat this argument by undermining the tacit internalism about qualia that was built into the first appeal to phenomenology. If the WRT is true there simply is no mental paint. It follows that my phenomenologically based belief that my experience is suffused with mental paint (in the form of detached qualia) is wrong.

My rejection of the belief thesis is based on this phenomenologically based belief. Therefore the truth of the WRT would seem to strengthen the belief thesis by defeating my defeater for the belief thesis. Thus, the belief thesis (and introspectionism) are epistemically in the clear. My admission that appeals to phenomenology cannot decide between the mental-paint and the physical-paint stories (see Section 7.4.2) appears to provide additional support for this argument. Hence, the persuasiveness of the supposedly obvious deliverance of phenomenology that our mental life is suffused with mental paint in the form of detached qualia is further weakened.

But this line of reasoning suffers from the same problem as the argument discussed in the previous section, viz., a confusion between issues pertaining to the nature of qualia with issues concerning the having of qualia. Whatever force the first appeal to phenomenology may have as an objection to the belief thesis belongs to it solely in virtue of its establishing that qualia can be had in a manner that is detached from corresponding qualitative beliefs. The further claim that the qualia that are had in this manner are internal mental properties plays no role at all in the argument against the belief thesis. We might as well drop this further claim about the nature of qualia, thereby acknowledging the inconclusiveness of the appeal of phenomenology inasmuch as the question of the nature of qualia is concerned.

The appeal to phenomenology is inconclusive only with respect to the nature-problem. This limitation in no way impairs its capacity to settle disputes about how qualia are had. Once it is understood which functions the appeal to phenomenology can and cannot play, one sees that the introduction of the relocation$_w$ move leaves its force against the belief thesis undiminished. I conclude that the introduction of wholeheartedly relocated qualia leaves the problematic status of the belief thesis unchanged.

When I first argued that the belief thesis has to be given up, the mental-paint thesis was tacitly presupposed. Now I argued that the introduction of the WRT does not change this verdict. Things do not look good for the belief thesis: no matter how the dispute about the WRT is resolved, the belief thesis fails. Inasmuch as I am concerned to refute the belief thesis I need not worry whether the paint is mental or physical. Thus, I can calmly face the possibility of a victory of the WRT. Nothing hangs on it as far as the belief thesis is concerned.

7.6.3 Wholehearted Relocation and the Existence of Qualia

In the course of the discussion of the WRT, I was forced to admit that the appeal to phenomenology cannot decide the issue of the location of qualia. For I had to acknowledge that the mental-paint scenario and the physical-paint scenario are

phenomenologically indistinguishable. What follows from this reversal? I will argue that, as far as the interests of the qualophile are concerned, the shift from the mental-paint scenario to the physical-paint story leaves everything as it was.

But a critic may feel that my professed equanimity is completely inappropriate. The critic might want to hold this view, even if he is willing to grant the point that the truth of the WRT does not weaken the case against the belief thesis. The critic's worry is a much deeper one. He might say this: By countenancing the possibility that phenomenal qualities are properties of nonmental external objects, you have effectively deprived yourself of your subject matter. To give up on the mentality of phenomenal qualities is to give up the claim that there are such things as qualia. Thus, it may seem to the critic that I have won the unimportant battle against the belief thesis at the expense of losing the war for the recognition of qualia as an important problem in the philosophy of mind.

The relocation move has, indeed, been employed as a strategy by those who want to rid the philosophy of mind of the troublesome problem of qualia. This strategy has, for example, been used by philosophers who attempt to immunize functionalism against the objection that it cannot account for the qualitative nature of pain. They suggest that the eliminativist and the reductionist strategies do not exhaust all the possibilities:

> We suggest a third strategy. We shall argue that, even if pain-qualia are essential for pain and even if pain-qualia cannot be functionally analyzed, that's no problem for Functionalism. It's no problem because, assuming they exist, qualia aren't psychological and, hence, don't fall within Functionalism's area of responsibility. (Graham and Stephens 1985: 73)

Graham and Stephens turn what seemed like a troublesome consequence of the WRT into an ingenious way of dealing with qualia-related problems for functionalism. But I cannot avail myself of this trick. For, naturally, I cannot hope to cast light on the nature of qualia by simply declaring them out of bounds. It may be a good policy to use this ploy to defend a narrowly conceived form of functionalism about the mental. If what you are after is to develop a theory of the mind, it makes sense to first purge the intended domain of your theory of all imposters. That is, it will make sense to first expose and then kick out all the purportedly mental items that are not genuinely mental. If qualia are such imposters, then good riddance. But if what you are after is an explanation of phenomenal consciousness, you will have to follow the qualia around wherever they may take you. If that is what you are interested in, it would be a mistake to insist at the outset that you will confine your study to the philosophy of mind, or to the philosophy of matter, or to epistemology, or to whatnot. Instead, you will study the qualia wherever they may happen to be located with whatever methods are

appropriate to the task. It is this latter project that I pursue. Therefore, I do not see the WRT *per se* as either posing a threat or as holding a promise for my enterprise. If the WRT is false and qualia are mental, then I will do philosophy of mind. If the WRT is true and qualia are nonmental, then I will have to do more than merely philosophy of mind. Shifting the qualia around does not make them go away. As long as the qualia exist somewhere, I will not be deprived of my subject matter. For this reason I face the outcome of the debate about the proper location of qualia with equanimity. I can do what I want to do no matter how the debate over the location of phenomenal properties turns out.

7.7 Summary

In this chapter we turned from the having-problem to the nature-problem. We listed four and discussed three ways in which to deny the traditional claim that qualia are nonphysical properties—three ways of deMenting qualia. The absence thesis denies that qualia are purely Mental because it denies that there are qualia. On a more plausible construal, the absence thesis reveals itself as simply another statement of the WRT. According to this view, qualia are properties of the external physical objects we perceive. But the seemingly powerful argument that favors the WRT over the HRT—the second appeal to phenomenology—does not keep what it promises. The phenomenological battle over the location of qualia has ended with a draw. Both versions of the relocation thesis receive powerful support from the consideration that there are no Mental qualia bearers. Relocation$_{w\ or\ h}$ presents itself as the only alternative to elimination of qualia. The WRT and the HRT each seem to face their share of problems. The problem of inflation confronts the WRT. The colored-brain thesis runs up against the obvious fact that the brain does not instantiate the qualia its owner experiences. But the main arguments against the relocation of qualia were postponed until Chapter Eight and Chapter Nine.

The latter part of the chapter was devoted to an assessment of the impact of the WRT on my own project. The conclusion I want to draw is this: the truth of the WRT leaves everything as it is. Locating qualia externally does not provide the opponent of the belief thesis with new objections against it, nor does it help the proponent of the belief thesis to fend off the old objections. Thus, the *status quo* is preserved and we are still justified in rejecting the belief thesis. This verdict also extends to introspectionism inasmuch as it depends on the belief thesis.[103] The details of the argument against the belief thesis (and, by implication, against introspectionism) do not turn on the truth or falsity of the WRT. The same is true of the project as a whole. The qualophile need not fear that the ex-

ternalization of qualia dooms the project of explaining consciousness in terms of qualia. Holding fast to the mental-paint picture of qualia is not a prerequisite to assigning them a crucial role in the philosophy of consciousness.

Up to this point we have encountered no (seriously) untoward consequences of the relocation$_w$ of qualia. This is encouraging for two reasons. First, it is in keeping with the neutrality thesis (see Section 2.2.2.3). Second, it counts in favor of an account of some nonbasic phenomenon (like consciousness) if the account is neutral with respect to more basic metaphysical questions. Regrettably, this peaceful coexistence of my project with the WRT will come to an end at the close of the next chapter. For there it will emerge that I am unable to comprehend the extant accounts of how relocated qualia—wholeheartedly as well as half-heartedly relocated qualia—are had. Nor do I know how to provide such an account of my own. To deny the WRT is therefore the only option I have, if I am to continue with the project of accounting for consciousness in terms of the having of qualia.

CHAPTER EIGHT

Having Relocated Qualia

It is all very well to affect a liberal air about the nature and location of qualia. But to allow for the existence and the exemplification of relocated celluloid qualia is not all it takes to account for qualitative consciousness. Otherwise we would have to assume that the lucky tomato, freshly painted with a new coat of red qualia, now has a phenomenal consciousness of its own. But surely, the tomato's inner light did not begin to shine any brighter by having its surface decorated with wholeheartedly relocated qualia. From the 'tomato's point of view' the relocationist's maneuver changes nothing: being a tomato with qualia is the same as being a tomato without qualia, viz., like nothing. Coloring the tomato with physical paint (relocation$_w$) or the neural state that is the perception of the tomato with mental paint (relocation$_h$) is only the first step towards an account of qualitative consciousness. For there to be consciousness, qualia need to be *had* in the right kind of way. But neither the tomato nor the neural state have them in this manner. If we are to be conscious we must have the qualia in the relevant way. What does it take to have the relocationist's qualia? How does one make them one's own? How can one manage to be conscious in a world of relocated celluloid qualia understood as physical or mental paint?

Scattered throughout the earlier discussion of the having-problem are pointers to the final installment of the anti-introspectionist argument. This concluding piece of the argument had to await the advent of the relocation theses. In turning to the nature-problem (in Chapter Seven), we have encountered the relocation theses. It is therefore time to turn back to the having-problem. The question we have to ask is this: Can (and does) introspecting or representing a quale let you have it in the right way (where having the quale in the right way is a matter of there being something that it is like to be you in virtue of your having the quale in this way)?

In the discussion that follows I focus on the more general case of representation without limiting the discussion to introspective forms of representation. This allows me to broaden the scope of the discussion—specifically, it allows me to address the anti-introspectionist representationalist position defended by Dretske and Tye—without sacrificing anything that is relevant to the special case of in-

trospectionism. When viewed from a sufficient distance, the differences between introspectionism and representationalism shrink to insignificance. On both accounts qualitative consciousness arises when a qualia-bearing entity (whatever its precise nature may be) is made the object of representation (whatever its precise nature may be). On both accounts, then, representation of qualia is at the heart of qualitative consciousness.[104]

As I see it, the evidence suggests that representation is not what lets one have one's qualia. My main complaint against a representation-based account of qualitative consciousness boils down to the claim that the account is ultimately unilluminating. It does not let us see how the representational activity is supposed to result in phenomenal feel. It does not help us understand why mobilizing representations (of the right sort in the right circumstances) results in its being like something to do so. This, together with my previous attempt to undermine the powerful intuition that (introspective) representation must hold the key to consciousness (Chapter Five and Chapter Six), plus my belief that an alternative account of how we have our qualia can be had (Chapter Ten), persuades me to reject introspectionism and, more generally, any representation-based account of qualitative consciousness.

8.1 Complicating the Picture: Four Representation-Based Theories

The question: Does representing one's relocated qualia let one have them? seems straightforward and unequivocal. But this impression quickly fades in the light of further reflection. Let us assume that we have an intuitive grasp of what it is to *have* a quale. Even so the question remains murky. Two of the other notions figuring prominently in this question require clarification. We need to know more about the roles that *representation* and *relocation* play in this context. By disambiguating these notions in different ways, we arrive at a number of non-equivalent questions about the relationship between representation and the having of relocated qualia. Thus we find many questions where there seemed to be only one. Each of the resulting questions is more precise than the original question. Regrettably, this gain in precision is not matched by a corresponding decrease in difficulty.

It has often been pointed out that the term 'representation' is ambiguous between the act, the content, and the object of a mental state. (See, e.g., Twardowski (1977)) The ambiguity that concerns us corresponds roughly to that between act and object. The first role that representations might play is that of qualia-bearers. Representations might be the internal objects that exemplify a person's qualia (that the person may or may not have at the time). The second

role that representations might play is that of qualia revealers. Representations might be the acts that let a person have their qualia (that are exemplified by a bearer inside or outside of the person). Illustrations of this ambiguity are not hard to find. By way of an example you may want to consider the two phrases 'representational theory of perception' (as it is traditionally used to characterize a Lockean theory of perception, for example) and 'representational theory of consciousness' (as it is used by Dretske and Tye to characterize their own view). In the first phrase the term 'representation' refers to internal qualia-bearing objects. In the second phrase it refers to the representational activity. These two roles for representations—that of internal qualia-bearer and that of qualia revealer—are quite independent of each other. An account of what it is to have a quale can enlist representations in the one, or the other, or both, or in neither function.[105] An answer to the having-problem can be counted as representation-based, if it invokes representations in at least one of these two functions.

Now to the term 'relocation.' The different interpretations of the relocation maneuver have already been canvassed in the previous chapter. The relocationist$_h$ moves the qualia out of the Mind and into the brain. The relocationist$_w$ pushes them all the way out onto the external physical object. What is involved in having relocated qualia will depend on the form this exile takes.

The two versions of the relocation thesis line up neatly with the two stances one can take towards representations in their role as internal, qualia-bearing objects. If we opt for relocation$_h$ and work with mental paint, then there will be some internal object that exemplifies those qualia. These internal objects play the role of representations understood as qualia-bearers. If we opt for relocation$_w$ and work with physical paint, then it is not necessary to postulate an extra set of internal representational objects to fill the role of qualia-bearers—the external objects themselves can play the part just fine. So we end up with two possibilities: either we have wholeheartedly relocated qualia without qualia-bearing internal representations, or we have halfheartedly relocated qualia in the brain with representations functioning as bearers of those qualia.

This alignment between the two versions of relocation (halfhearted and wholehearted) and the two stances on representational objects (acceptance and rejection) does not touch upon the question of representational acts. Quite independently of the question whether representations are involved in bearing our qualia, we still face the question whether representations play a role in revealing these qualia to us. Again we have two options: the disclosure of qualia relies on some form of representational activity on our part, or it proceeds in some other way that does not involve acts of representation.

We started with a single question: Does representing one's relocated qualia let one have them? Reflection on the different interpretations of the notions of *relo-*

cation and *representation* shows that the situation is more complicated than this simple question suggests. Combined, the two distinctions drawn above yield four different ways for representations to be involved in the having of relocated qualia. The rows in the following table stand for the location and bearers of qualia; the columns stand for the ways in which these qualia are revealed:

Table 2. Preliminary chart of representation-based accounts

	Qualia revealed by acts of representation	Qualia not revealed by acts of representation
Relocation$_h$: qualia borne by representations in the brain	I	II
Relocation$_w$: qualia borne by external objects	III	IV

Fortunately we need not pay attention to all the cells in this table. Cell (IV) falls outside the scope of a representational account—for it describes a situation in which qualia are neither borne nor revealed by representations. I suggest that we also skip cell (II). Since our primary interest concerns the question whether and how the act of representing qualia can let us have them, we need not pursue this possibility here.[106] This leaves us with cells (I) and (III).

But just when things begin to look simpler we encounter another complication. Those who enlist acts of representation to disclose our qualia do so in quite different ways. As we have seen, introspectionism offers a hierarchical account that insists that higher-order representation is necessary for the having of qualia. On the other side we find nonhierarchical theories that do not rely on introspection to let us have our qualia. The representationalist (or intentionalist) positions of Dretske and Tye are cases in point. According to these authors, a single level or representation is sufficient for one to have one's qualia. Thus we are back at a table with four cells. The rows are taken over unchanged from the previous table, but the columns differ.

Table 3. Final chart of representation-based accounts

	Qualia revealed by higher-order representation	Qualia revealed by first-order representation[107]
Relocation$_h$: qualia borne by representations in the brain	1. Pollock Rosenthal	2. Maloney
Relocation$_w$: qualia borne by external objects	3. Armstrong Lycan	4. Dretske Tye

This time the contrast is between two distinct kinds of representational activity involved in the disclosure of qualia. That is, the columns represent what we might call an inner-representationalist dispute. The question is not whether representational acts are needed to let us have our qualia; the question is what sort of representational action achieves this feat.

Note that the cells (I) and (III) of Table 2 do not correspond precisely to cells (1) and (3) of Table 3. The position represented by cell (I) of Table 2 is ambiguous between the positions represented by cells (1) and (2) of Table 3. Parallel remarks hold for cell (III) and cells (3) and (4).

There are, then, four representation-based accounts of what it is to have relocated celluloid qualia. I will now argue that none of the four versions is satisfactory. This is so because not one of these accounts makes it intelligible why relating oneself to one's qualia in the specified manner results in its being like something for one to do so. I end this chapter with an impressionistic sketch of the position defended in Chapter Ten. I present this preview at this juncture because I hope that an appreciation of the problems faced by the four representation-based accounts of the having-problem might incline a reader to take seriously a position that might otherwise have struck her as gratuitously bizarre.

8.2 Cell (1): Internal Qualia with HO Representation

In Chapter Six I presented a extended critique of Pollock's account of qualia and how they are had. The details of theories belonging into cell (1) vary. But I trust that the sorts of arguments enlisted against the Pollockian view will—once the necessary changes have been made—also have some force against the other variants of this general form of introspectionism. The need for a lengthy discussion of Rosenthal's HOT theory of consciousness is thus obviated. But I will address one powerful objection that a number of critics have leveled against HOR accounts of both the cell (1) and the cell (3) variety.

8.2.1 The Problem of the Rock

At the heart of Rosenthal's thinking about consciousness is a quite general principle about explanation: intrinsicalness impedes informative explanation; being relational is a necessary condition upon a property's having an illuminating explanation. Applied to consciousness this yields the result that "only if we regard being conscious as a relational property…can we explain what such consciousness consists in." (Rosenthal fc.: Section III) If consciousness is taken to be an intrinsic property of mental states, the route to a nontrivial, informative explanation of the phenomenon is blocked. Were we to succumb to the Cartesian temp-

tation to view consciousness as intrinsic (and therefore as simple and unanalyzable) we would have to

> acquiesce in our inability to give any genuine explanation of consciousness at all; at best we can render the phenomenon more intelligible by tracing the merely conceptual connections that hold among such cognate notions as mind, consciousness, subjective, viewpoint, first-person perspective, and self. (Rosenthal fc.: Section III)[108]

Which is it going to be: a genuine, informative explanation or a mere tracing of conceptual connections? The way this contrast is phrased makes it patently clear that a genuine explanation is what we want. Were we to discover that a tracing of conceptual connections is the best we can get we should feel intellectually defeated. What Rosenthal tells us is that holding fast to the view that consciousness is an intrinsic property is a guarantee of intellectual defeat. Thus the relational nature of consciousness is revealed as a condition of the possibility of an intellectually satisfying understanding of consciousness. Guided by these weighty assumptions about the nature of explanation, Rosenthal makes the search for a relational construal of consciousness into the centerpiece of his project. As we have seen in Chapter Four, he takes himself to have found the crucial relation, viz., being accompanied (in a suitably direct way) by a higher-order thought that makes the lower-level mental state conscious.

Functioning as a lightning rod of sorts, Rosenthal's bold proposal has attracted critical flashes from a variety of thinkers who do not find much to agree on otherwise. Here I present Alvin Goldman's version of a by now classical objection to HOT or, more generally, HOR approaches to consciousness. Goldman puts the worry into simple but powerful words: "A rock does not become conscious when someone has a belief about it. Why should a first-order psychological state become conscious simply by having a belief about it?" (Goldman 1993a: 366) Elaborating on the theme he writes:

> the underlying idea here is puzzling. How could possession of a meta-state confer subjectivity or feeling on a lower-level state that did not otherwise possess it? Why would being an intentional object or referent of a meta-state confer consciousness on a first-order state? (Goldman 1993a: 366)

Let us use the label 'the rock' as a shorthand way of referring to this objection and the problem it raises.[109] It is a problem with which HO theorists have often been confronted. So it comes as no surprise that Rosenthal has a ready answer to this challenge.

8.2.2 Mediate and Immediate HO Thoughts

Before presenting Rosenthal's specific reply, I want to briefly present another thought of Rosenthal's that bears on this issue. Rosenthal acknowledges that higher-order thoughts do *not* invariably lead to consciousness: "One can be conscious of being in a mental state even when we would not count that state as being a conscious state." (Rosenthal fc.: Section IV) This concession may spark a hope in the critic of HO accounts: if higher-order representation does not always issue in consciousness then, perhaps, it can be argued (from *within* the HO position) that it never does. This may just be the proverbial chink in the enemy's armor. But Rosenthal senses no danger from this direction. The cases he means to exclude are those in which the higher-order thought is, in some sense, insufficiently direct. Think of the sort of case in which psychological theorizing leads you to entertain a belief about one of your mental states that you cannot access otherwise. What this case suggests is a restriction upon the relevant higher-order thoughts: they must be "suitably unmediated" (Rosenthal fc.: Section IV) or "suitably immediate." (Rosenthal 1991: 31)

The addition of this rider satisfies the HO theorist that the chink in his armor is closed—for one cannot have a *suitably immediate* thought about an unconscious mental state without thereby making it conscious. But the critic of the hierarchical account is not silenced that easily. What he fails to understand is how *any* thought, be it direct or indirect, could make the unconscious conscious. This is the worry articulated in the problem of the rock. It is, of course, a worry that the critic brings to the HO theory 'from the outside'—a worry that he cannot expect to be shared by philosophers of Rosenthal's persuasion. It is one of these cases where what seems obvious to one group strikes the other group as incomprehensible.

8.2.3 The Official Reply to the Rock

This brings us back to Rosenthal's official answer to the problem of the rock. In stating his reply, Rosenthal uses the terms 'transitive consciousness' and 'intransitive consciousness.' The distinction turns on the notion of intentional directedness. Consciousness that has this directedness Rosenthal calls 'transitive'; consciousness that lacks this directedness he calls 'intransitive.' Only a subject can be directed at, or conscious of, an object; the subject's states cannot be said to have this characteristic. Thus, transitive consciousness is reserved for subjects, and state consciousness is intransitive. With this terminological clarification, we are ready for Rosenthal's reply to the problem of the rock:

> We must distinguish two ways of understanding the present proposal. Being transitively conscious of a mental state does in a sense make it intransitively conscious. But that is not because being conscious of a mental state causes that state to have the property of being intransitively conscious; rather, it is because a mental state's being intransitively conscious simply consists in one's being transitively conscious of it. The mistake here is to suppose that a state's being intransitively conscious is an intrinsic property of that state. If it were, then being intransitively conscious could not consist in one's being transitively conscious of being in that state unless being thus conscious induced a change in that state's intrinsic properties. This objection is at bottom just a disguised version of the doctrine that being intransitively conscious is an intrinsic property. (Rosenthal fc.: Section IV)

What Rosenthal provides in this passage is not so much an argument as a diagnosis conjoined with a prescription. Resistance against the HO account of state consciousness is grounded in a tacit commitment to the intrinsicalness of consciousness. Given this view of consciousness, it is mysterious how consciousness arises *in* a state merely because some other mental state represents it. So much for the diagnosis. The remedy consists in reconceptualizing consciousness as a nonintrinsic, relational property. On such a view the HO account of consciousness loses its air of paradox.

8.2.4 The Rock and Intrinsicalism

To assess the diagnosis we have to take a closer look at Rosenthal's notion of an intrinsic property. "A property is intrinsic" he tells us "if something's having it does not consist, even in part, in that thing's bearing some relation to something else." (Rosenthal fc.: Section III) Is it true that those who are troubled by the problem of the rock must hold that the consciousness of a state is intrinsic to it in this sense? I think not.

The reason for doubting Rosenthal's diagnosis is grounded in the feature of state consciousness that I have variously referred to as the subject involvement of consciousness, or the *for-someone feature* of consciousness, or the fact that consciousness has built into it a direction upon an apprehending mind. That a given state is conscious means that having the state makes it be like something to be the owner of the state. In this way state consciousness points beyond the state toward an experiencing subject. And this relation to the experiencing subject is a necessary condition upon the state's being conscious. There cannot be a conscious state without a subject whose state it is. Thus, state consciousness is not intrinsic to the state in Rosenthal's sense. So it seems that we have to reject Rosenthal's diagnosis of the rock—the objection is not "just a disguised version of the doctrine that being intransitively conscious is an intrinsic property." (Rosenthal fc.: Section IV)

What, then, is the correct diagnosis of the rock? This question must be answered with some care. For my attempt to salvage the rock may appear like a mixed blessing. On the one hand, it seems that Rosenthal must retract his claim that the problem of the rock is not so much an argument against his view as merely a restatement of a contrary doctrine—the doctrine of intrinsicalism. On the other hand, it may seem that my 'critique' of Rosenthal's diagnosis ends up supporting his cause. For, if I am right and the proponents of the rock are not intrinsicalists—well, then they must hold a relational view of consciousness! And once one embraces a relational account of state consciousness the worry expressed in the rock simply does not arise. So it may seem that my argument backfires. Rather than offering support to the advocates of the rock, it appears to undermine their position and to suggest that they switch allegiances and embrace Rosenthal's representational account of state consciousness. In short, it looks as if I unwittingly played into Rosenthal's hands.

It is always a pleasure to see someone hoist by their own petard. But in this case the pleasure may have been premature. First, consider the following question: Is the rejection of intrinsicalism about state consciousness tantamount to accepting Rosenthal's relational account of consciousness? As the question suggests, the answer is No. Rosenthal advocates a purely relational theory of state consciousness. According to this view, the intrinsic nature of a mental state is independent of what we may call its 'consciousness value,' to use a self-explanatory neologism. Neither the current consciousness value nor changes of this value are in any way reflected in the inner nature of the mental state. The inner nature of a mental state is not constitutive of, indicative of, or causally involved in, the state's being conscious or unconscious. The external relations of a state are the sole determinant of this fact.

The denial of intrinsicalism is a much weaker doctrine. It merely denies that the inner nature of a mental state is the sole determinant of its status as conscious or unconscious. While this denial acknowledges the role played by an external, relational component, it leaves room for the possibility that the internal condition of a mental state plays a role in constituting, or indicating, or causally producing the state's consciousness value. Thus, the denial of intrinsicalism does not entail Rosenthal's purely relational theory of consciousness. So the argument—you deny that friends of the rock are intrinsicalists, therefore, you make them into relationalists à la Rosenthal—fails. This blocks the quick move from the denial of intrinsicalism to the demise of the problem of the rock.

An analogy may help clarify this dispute about the nature of state consciousness. Take a compass and a magnet. Place the magnet south of the compass. Now consider the following two properties of the compass. First, it is to the north of the magnet. Call this property 'northerly.' Second, it is caused to point

to the south by the magnet. Call this property 'southerly.' The magnet's location is the sole determinant of the compass's northerlyness. The inner makeup of the compass plays no part in this. Thus the compass's being northerly is a purely relational property. The property southerly is not purely relational. For if the needle—a part of the compass's inner makeup—does not point south the compass isn't southerly. And the property southerly is not intrinsic. For, if there is no magnet—a distinct existence, no less—influencing the needle, the compass isn't southerly.

The difference exhibited by the properties northerly and southerly mirrors the difference between the two notions of state consciousness we have encountered. For Rosenthal a mental state's being conscious is, in the relevant respect, analogous to a compass's being northerly. The innards of the mental state and the compass are alike in, respectively, being irrelevant for the former's being conscious and the latter's being northerly. The friends of the rock disagree. The property southerly is the relevant analog to state consciousness as they understand it. If the relevant internal conditions of the mental state and the compass are not satisfied, then the addition of external relations is incapable of making the former conscious or the latter southerly.

The compass analogy may help one to see that resisting an intrinsicalist account of something (southerlyness, for example) is not the same as embracing a purely relational theory of it. For in the case of southerlyness (unlike in the case of state consciousness!), it seems quite obvious that and why the latter course is hopeless. This suggests that attempting a purely relational theory of consciousness is no more promising than attempting to account of a compass's southerlyness without mentioning the orientation of its needle. Keep in mind, however, that this is just an illustrative analogy and not an argument against the purely relational account. It cannot settle a dispute. But perhaps it can help to bring out the nature of the dispute in fuller relief.

8.2.5 HOT: Analysis not Explanation

How much weight should be attached to this point? Perhaps the denial of intrinsicalism does not *entail* the purely relational theory—but it may still be the case that the purely relational theory is the only reasonable choice once intrinsicalism has been rejected. A satisfactory argument to the contrary would have to show that (i) there is at least one viable alternative to the purely relational theory of state consciousness and that (ii) this alternative is preferable to the purely relational theory. To show this is a large undertaking. Here I only broach one part of the argument in support of (ii). If it can be shown that the purely relational account is unsatisfactory, then it won't take much for the alternative account—

assuming that there is one—to be the more attractive of the two. To argue that there is such an alternative account that it is not altogether unbelievable and has certain virtues, is the task of Chapter Ten. The question at hand concerns the intrinsic attractiveness of the purely relational account.

This discussion opened with a presentation of Rosenthal's principles governing informative explanation (see Section 8.2.1). Genuine, informative explanation was contrasted with the tracing of mere conceptual connections. The promise held out to us was that in return for adopting a relational construal of consciousness we would be rewarded with a genuine, informative explanation of consciousness. The question we now face is this: Did Rosenthal uphold his end of the bargain and give us a genuine explanation rather than a 'mere conceptual analysis'? I will argue that, contrary to the advertisement, the HOT theory is best seen as a 'mere conceptual analysis.' Unlike Rosenthal, I see no fault in this—illuminating conceptual analyses are the best you can hope for in philosophy. My reservation concerns only the amount of light that this particular analysis sheds on the question of qualitative consciousness.

Stripped of all qualifications Rosenthal's account says that "a mental state is conscious just in case we are transitively conscious of it." (Rosenthal fc.: Section IV) Should this formula strike you as exuding the stale odor of conceptual analysis—rather than the robust scent of a genuine explanation—Rosenthal hastens to add that "this proposal is a substantive hypothesis about what intransitive consciousness is, not a recommendation about how to conceive of it." (Rosenthal fc.: fn. 29) So there can be no doubt as to Rosenthal's intentions—he wants to be understood as providing a substantive, i.e., I assume, an empirically grounded and empirically relevant conjecture about the human mind. Substantive hypotheses require substantive support. Accordingly, we expect to find Rosenthal marshaling empirical observations, both familiar and scientific, in support of his conjecture. But we find nothing of the kind. Rosenthal's paper is a paradigm of philosophical argumentation—something quite unlike anything a biologist or psychologist of consciousness would ever produce. His arguments proceed from the mixture of general philosophical principles and pretheoretic intuitions that is so characteristic of philosophical dialectic. To support this claim with an exhaustive study of Rosenthal's arguments would be exhausting—for his paper is a paradigm of philosophy not just as measured by the quality of argumentation but also by the quantity of argument. So I will limit myself to just a few remarks on this issue.

Let's begin by looking at the distinction between transitive and intransitive consciousness. This distinction lies at the heart of Rosenthal's HO account of state consciousness. It is what allows him to explain consciousness (intransitive) in terms of consciousness (transitive) without falling into a circle and without

tautologizing. What underwrites this crucial distinction? A simple appeal to common usage is all it takes. He observes that some of our uses of the term 'consciousness' take a direct object, others don't. And that is what grounds the distinction between transitive and intransitive consciousness. (See Rosenthal fc.: Section IV) So this is one case where an appeal to common linguistic intuitions gets us launched.

Once this distinction is in place it has to be put to work. How shall we proceed? Commonsense and a sense of naturalness point the way. The HO model that explains intransitive consciousness in terms of transitive consciousness is the "natural suggestion"; it "squares reasonably well with our pretheoretical intuitions"; and "captures our commonsense intuitions." (Rosenthal fc.: Section IV) Again it is the appeal to ordinary intuitions that gets Rosenthal off the ground.[110]

Finally, look once more at the way in which Rosenthal counters the objection of the rock. Replying to his opponents who stubbornly insist that a lower-level state can be made conscious by a higher-level state only if the latter brings about an intrinsic change in the former, Rosenthal writes: "A mental state's being intransitively conscious *simply consists* in one's being transitively conscious of it." (Rosenthal fc.: Section IV, italics added) The phrase 'simply consists' suggests that the right way to understand the quoted sentence is to take it as expressing something stronger than a purely empirical result. What it seems to say is that careful reflection on our pretheoretic concepts should suffice to discover the relational nature of state consciousness. This reading also recommends itself in the light of Rosenthal's earlier appeals to commonsense, pretheoretical intuition, etc. Again we are left with the impression that Rosenthal's central result is arrived at by tracing out concepts and not by some more substantive method.

It is instructive to compare Rosenthal's reply to the problem of the rock with Lycan's treatment of this objection. Lycan, never one to beat around the bush, states his reply to the problem of the rock with disarming frankness. Consideration of his reply will help crystallize the suspicion evoked by Rosenthal's more guarded approach. Lycan writes:

> What is so special about physical states of that certain sort, that consciousness *of* them makes them "conscious"? That they are themselves mental...It seems psychological states are called "conscious" states when we are conscious of them, but nonpsychological things are not. (Lycan 1995b: 3)

So it seems indeed. But this observation will not satisfy one who is troubled by the problem of the rock. What she wants to know is *why* mental states work in this way. What is it about a mental state that enables it to become conscious by acquiring a 'converse intentional property' (Chisholm's term)? Lycan addresses

this query, albeit in a most unexpected manner. Expanding on the answer we just saw, he writes:

> Given the reality of the distinction between states we are aware of being in and states we are not aware of being in, the only remaining question is that of why the *word* "conscious" is thus dragged in as an adjective to mark it. My bet is that there is a grammatical answer. Maybe it is a transferred epithet: We begin with the adverbial form, as in "consciously thought" or "consciously felt," and when we make the verb into a noun the adverb automatically becomes an adjective—as in the move from "meditatively sipped" to "took a meditative sip." That is fairly plausible; at any rate it is the best I can do for now. (Lycan 1995b: 3-4)

The first installment of Lycan's answer left the critic unsatisfied. The second installment is liable to render her speechless. A mind in the grip of the problem of the rock reels when confronted with the suggestion that the 'mysterious power' of internal monitoring to produce (state) consciousness boils down to a transferred epithet not recognized as such.

However, our present concern is not how well Lycan's reply succeeds in dealing with the objection of the rock. I presented his answer to the rock only to illustrate a position—a position in many ways very closely related to Rosenthal's—according to which the question of state *consciousness* does not demand a substantive answer. Lycan grants that there is a substantive issue in the background: some mental states are, and others are not, accompanied by higher-order monitoring. This the proponents of the rock will gladly acknowledge. But this is also where the agreement ends. Lycan holds that any further questions about consciousness are really questions about 'consciousness.' It is all just a matter of how we came to use the term in this way. The proponents of the rock disagree. They hold that the further question, the question how higher-order representation engenders state consciousness, is a (or, perhaps, *the*) substantive question. On this point Rosenthal appears to align himself with Lycan's critics by insisting that the nature of state consciousness poses substantive questions. The disagreement between Rosenthal and those who present the rock objection is 'merely' about the details of the correct account of state consciousness. What I have suggested above, and what I hope the comparison with Lycan's extreme view has helped to bring into clearer focus, is the possibility that Rosenthal's account of state consciousness is less substantive than he would have us believe. I do not mean to imply that Rosenthal follows Lycan in holding that the issue is merely grammatical. But I do think that Rosenthal's theory of state consciousness is better understood as a product of traditional philosophical analysis than as a piece of scientific theorizing.

What I am proposing, then, is this: Rosenthal's HOT account of state consciousness is best viewed as precisely the sort of thing we get from "tracing the merely conceptual connections that hold among [a number of] cognate notions…" (Rosenthal fc.: Section III) In thus reclassifying Rosenthal's account of consciousness I do not mean to criticize it. Unlike Rosenthal, I do not regard the fruits of conceptual clarification as the poor relations of genuine explanations. Successful conceptual analysis can be quite as illuminating as scientific explanation. But just as some explanations do not explain, some analyses do not shed light. It is on this count that Rosenthal's attempt to elucidate consciousness is wanting, as is evidenced by the problem of the rock.

8.2.6 Assessing the Rock

Rosenthal dismissed the problem of the rock. I still want to insist on its importance. Therefore, a brief reminder of the reasons for resisting Rosenthal's way of disposing of the rock is in order.

Rosenthal's account of state consciousness does not live up to its billing—it is not the substantive explanation of state consciousness we were promised. It is better understood as resulting from analyzing and sharpening our pretheoretic notion of consciousness. But it is none the worse for *that*. Rather than concern ourselves with the status of the account—Is it a scientific theory or is it 'merely' a philosophical analysis?—we should ask whether it does justice to the phenomenon of consciousness. I argued that Rosenthal's thoughts about state consciousness—and the problem of the rock in particular—are informed by a false dichotomy. It is not true that we must choose between the purely relational position he advocates and the intrinsicalism of which he accuses his opponents. The existence of intermediate positions opens up the possibility that the objection of the rock is not grounded in intrinsicalism. I have argued that reflection on the nature of state consciousness suggests just that. State consciousness has a relational component, for a conscious state is always conscious *for someone*. So the relation to the apprehending subject is ineradicable. But acknowledging this relational aspect of state consciousness falls far short of accepting Rosenthal's purely relational account. If this point is granted, Rosenthal's principal reply to the rock collapses. On the grounds that the rock is merely a disguised statement of intrinsicalism he argued that it is not so much an objection against the HO account as simply a denial of the account. I hope to have shown that this is not true. The rock poses a genuine problem.

How great is the damage wrought by its impact? To better answer this question let us bring into focus the concerns that motivate the proponents of the rock. If intrinsicalism about state consciousness isn't the driving force behind the

rock, what is? To raise the problem one need not be committed to the intrinsicalness, ultimate simplicity, and in-principle unanalyzability of consciousness. On the contrary—one may be quite open to the idea that the relations into which a state enters are among the constitutive factors of the state's consciousness. But such openness is compatible with doubts about the purely relational account of consciousness. What motivates the proponent of the rock is the worry that the relation *being the target of a higher-order representation* is the wrong relation. The worry is fueled by one's inability to comprehend how entering into this relation is supposed to promote an unconscious state to consciousness. Those who raise this objection aren't just begging the question against Rosenthal by simply restating their intrinsicalist creed. They attempt to draw attention to the inadequacy of a purely relational account of state consciousness. Rosenthal's self-assured claim that a state's being conscious 'simply consists' in its having a higher-order thought directed at it strikes them as incomprehensible. We have not been given a mechanical model or a tracing of conceptual connections that begins to make intelligible how this hierarchical arrangement could eventuate in a state that it is like something to have.

8.3 Cell (2): Internal Qualia without HO Representation

We are now looking at cell (2) of Table 3 on page 184. That is, we are looking at a position that combines the relocation$_h$ of qualia with a nonhierarchical account of how these qualia are had. This is precisely the sort of position that J. Christopher Maloney has defended. His view is best characterized as a 'representational use theory of consciousness.' He summarizes it as follows:

> To be conscious is...simply, but *actually, to represent.* To be conscious in a way peculiar to a kind of conscious agent is *literally,* not metaphorically, *to use* a system of representation characteristically used by that kind cognitively to represent. (Maloney 1985: 26)

Applied to our own case, the theory holds that "what it is like to be a human is to use the system of representation naturally used by, and perhaps peculiar to, humans." (Maloney 1985: 30) In the present intellectual climate the idea that thought employs a representational system is often not distinguished from the idea that thought is a matter of computation. Given this background, Maloney's theory of qualitative consciousness does seem to have all the virtues and vices of computationalism at its most brazen: it is clear, simple, and unbelievable. For doesn't it seem positively miraculous that the manipulation of formal, computational symbols (for that is what representations are) should eventuate in the light of qualitative consciousness coming on?[111]

Maloney has foreseen this objection. He points out that actual computation with formal symbols involves more than merely abstract symbol manipulation: the symbols as actually used by any given system are *embodied* in one way or another. That is, symbols in use have not merely their formal properties; they also have physical properties on account of their physical instantiation in a physical being that uses them. And Maloney argues that these physical properties of representations are the proper locus of phenomenal properties: "The theory…identifies the qualia of consciousness with the kinds of physical structures realizing cognitive events." (Maloney 1985: 27) Adopting the popular thesis that the system of mental representations is best viewed as a language-like structure—the language of thought, or Mentalese—Maloney puts the same point in different words when he writes: "It is the physical, structural properties of an agent's Mentalese token placing the token in its proper Mentalese type that are the agent's phenomenological qualia." (Maloney 1985: 31) Maloney's move is a tempting one for a computationalist in search of more elbow room. By identifying qualia with the physical aspects of representations, Maloney can now avail himself of the explanatory resources of orthodox computationalism as well as those of physicalism.[112]

While I think that Maloney has taken a step in the right direction, I still find myself unsatisfied. His representation-based theory accounts for the instantiation of qualia. It does that by making the qualia (of human beings) into properties of neural state tokens. Thus Maloney endorses a version of the halfhearted relocation thesis (HRT)—the colored-brain thesis. In the light of the problems besetting the colored-brain thesis (see Section 7.5 above, where the colored-brain problem is discussed) one cannot help but wonder whether the properties that Maloney wants to attribute to the neurophysiologically instantiated tokens of mentalese are genuine, silver-screen qualia or merely some poor relatives of theirs—perhaps something like their "neurophysiological correlates." This is the term that Francis Crick uses in his more guarded moments to describe the properties that future neurophysiology will discover. (Crick 1994: 9-10) When pressed hard, Crick appears to acknowledge the possibility that qualia proper might forever be beyond the scope of neuroscience! (See Crick (1994: 258)) Crick's move avoids the problems of the colored-brain thesis. But it also voids whatever explanatory purchase neurophysiology may have on qualia. Since Maloney denies the distinction between qualia and their neurophysiological correlates by identifying them, it appears that he is committed to the view that our brains exemplify the phenomenal colors as we experience them. This view is also supported by the fact that Maloney does not require that qualia be introspectively represented in order to be had. He does establish an intimate link between qualia and representational systems by making qualia into features of the imple-

menting hardware. But unlike regular celluloid qualia, Maloney's qualia need not be represented in order to be had. This suggests that Maloney works with silver-screen qualia. I side with the proponents of physical paint in holding that this is an unacceptable conclusion. Dretske is right: casual inspection shows that qualia are nowhere to be found in the head (see Section 7.5.2 above).

If one wants to hold fast the idea that qualia are mental paint, there are only two ways around this problem. First, one might try develop a new model of how silver-screen qualia inhere in the brain that sidesteps the colored brain problem. The model has to allow us to (i) say that the qualia you have are in your brain and (ii) deny that your brain exemplifies the qualia you have (in the way in which it exemplifies greyness, for example). I attempt to develop this idea in Chapter Ten. But Maloney does not pursue this line of thought. Second, one might insist that one is not committed to silver-screen qualia after all. In support this contention one might argue that introspective representation may not be the only process that lets one have a celluloid quale one instantiates. If there exists an alternative 'having mechanism' then the denial of the claim that qualia are had by means of introspective representation does not entail a commitment to silver-screen qualia. In taking this approach one would broaden the notion of a celluloid quale. One would still insist that a celluloid quale is one that needs a having mechanism in order to be had. But one would no longer insist that this having mechanism be representational.

I think that Maloney's position can plausibly (and charitably) be interpret along these lines. His talk of the *use* of representations suggests that he has an alternative having mechanism in mind. And the switch to celluloid qualia has the enormous advantage that the colored brain problem never arises. Though Maloney's theory looks much better now we are still left with the having problem: How does one get to have the celluloid qualia that are properties of one's neural states? As yet the question is unanswered.

Remember Nagel's gourmet scientist who licks his subject's brain so as to learn whether the brain's taste matches the subject's taste experience. The discussion of Nagel's scientist served to make two points. First, your brain does not exemplify the qualia you *have*. Second, even if your brain did exemplify the qualia you have, its exemplifying them is no explanation of your having them. Maloney denies the first claim. This denial is controversial and interesting, but I will not discuss it here because this would lead us away from our main problem, viz., the having-problem. Even more controversial and more interesting is Maloney's denial of the second claim. Given one minor addition, Maloney is prepared to defend the view that Nagel rejects: your having certain qualia is explained by your brain's exemplifying them. The additional condition that Maloney insists on is this: the states of your brain that exemplify the qualia you

have must be the states that you token in having those qualia. So we get the conclusion that your brain's exemplifying the qualia you have does explain your having these qualia, given that the neural states exemplifying the qualia are the representations you token in having these qualia. Here is Maloney's account of how we have our qualia in his own words:

> It is the physical, structural properties of an agent's Mentalese token placing the token in its proper Mentalese type that are the agent's phenomenological qualia. To use a Mentalese token embodied as a token of a specified physical type is to experience the phenomenological qualia associated with the occurrence of the thought employing that Mentalese token. (Maloney 1985: 31)

This, then, is Maloney's theory of how we manage to have our qualia: to *use* a representation that exemplifies a given quale is to *have* the quale that the representation exemplifies.

At the heart of Maloney's representation-based theory of consciousness is the notion of representational use. In using a representation you make its qualia your own. But surely it is fair to ask: How does this work? How does representational use bring it about that a quale exemplified by a neural state of mine is thereby *had* by me? These questions are reminiscent of the questions raised by Rosenthal's purely relational account. Rosenthal told us that a state's consciousness simply consists in its being accompanied by the relevant HOT. Maloney tells us that a state's consciousness simply consists in the relevant use of a representational state. So the views differ considerably. Maloney tries to get by without the superstructure of higher-order representation that is the centerpiece of Rosenthal's purely relational account. But despite the diversity of the apparatus enlisted in the account of how our mental-paint qualia are had, the views are similar in the ways in which they fail to satisfy. Both would be unsatisfactory as explanations even if known to be true. Assume that Maloney is right and we know it. Would this answer all our questions about state consciousness? No, because we still would not understand how using representations in this way could result in there being something it is like to be the representation user. The situation is similar to the one we encountered when dealing with the views of Pollock and Rosenthal. The explanatory resources mobilized to account for the phenomenon of qualitative consciousness seem inadequate to the task. If one of these accounts were discovered to be true we still would not grasp how this could be so.[113]

In Maloney's version of representationalism the primitive notion of representational use is made to carry all of the explanatory load. The use of embodied, qualia-bearing tokens of Mentalese makes the user have the qualia. But how or why this should be so, we are not told. For all I know Maloney may be right

about what he says. But the fact, if it is a fact, that representational use makes one be qualitatively conscious strikes me more as an explanandum than as an explanans. Thus, I find myself incapable of accepting his theory as an explanation of how (halfheartedly) relocated qualia are *had*. I can conclude my animadversions against Maloney's representationalism in a manner that parallels my conclusions about HO accounts of qualitative consciousness. I find Maloney's representationalism unilluminating in the following sense. Just at the point where things get interesting we run out of explanations. We are left with a big, baffling fact instead of an intelligible account of that in virtue of which this fact is a fact.

8.4 Cell (3): External Qualia with HO Representation

The attempt to explain consciousness by combining wholeheartedly relocated celluloid qualia with the machinery of higher-order representation (cell (3) of Table 3 on page 184) is a classic among contemporary theories of qualitative consciousness. David Armstrong and, following his lead, William Lycan are the most visible thinkers in this camp. A remarkable feature of this position is the insistence that the HO representations involved in introspection are to be understood as inner sensing or perceiving—hence the label 'HOP.' This gives the theory a sharper definition; it also makes it more vulnerable. In assessing the import of relocated qualia for introspectionism, I pursue two main strategies. The first looks back to the earlier engagement with introspectionism. Picking up the thread from Chapter Five and Chapter Six, I ask how the old anti-introspectionist arguments affect the introspectionist doctrine after its transfer into a world of physical paint. The second line of inquiry exploits a specific feature of the Armstrong/Lycan approach. They construe introspecting as inner sensing. Relocated qualia are outer properties. I ask how the inner sense is supposed to get at the outer qualia.[114]

8.4.1 Introspectionism and Relocation

The current position is a version of introspectionism. That means that all the arguments directed against introspectionism in Chapter Five and Chapter Six also apply to it. The point of Chapter Five was that introspectionism is not well motivated. The point of Chapter Six was that there are good reasons for thinking that introspectionist accounts of consciousness are false. The one big difference between the theory we are considering now and the theories considered in Chapter Five and especially Chapter Six is that we are now dealing with wholeheartedly relocated qualia. So the relevant question we need to ask is this: Does the switch to the physical-paint scenario undermine or strengthen the anti-introspectionist

arguments set out in Chapter Five and Chapter Six? As best I can tell the arguments of Chapter Six remain largely unaffected by the switch from mental paint to physical paint. Had Pollock opted for the relocation$_w$ of Oscar's qualia, the argument from blindsight (Section 6.4.7) together with the supporting arguments presented there would still give us good reason to doubt that Oscar is conscious. The impact of the switch to mental paint on the arguments of Chapter Five is not so easily assessed. But fortunately this task has already been carried out in Chapter Seven. There I argued that the relocation$_w$ of qualia leaves everything as it is. First, it is not the case that the switch to physical paint affords a knock down argument against the belief thesis (Section 7.6.1). Second, it is not the case that the switch to physical paint reinstates the belief thesis by undermining the first appeal to phenomenology, which was its main defeater (see Section 7.6.2). So the relocation of qualia leaves untouched the case that was mounted against the belief thesis in Chapter Five. Given that the belief thesis is the main source of plausibility for introspectionism, the collapse of the belief thesis in the physical-paint scenario shows that introspectionism is unmotivated in the world of physical paint. So we have seen that the old arguments against introspectionism show that introspectionism is no better motivated and no more likely to be true in the new physical-paint scenario than it was in the old mental-paint scenario.

So much for the reassessment of the old arguments against introspectionism in the new setting of the WRT. The upshot is that the status of introspectionism remains unchanged. But the next section will show that the switch to physical paint affords us a new and more powerful reason to doubt the viability of an introspectionist account of consciousness is a world of wholeheartedly relocated qualia.

8.4.2 The Inner Sense and the External World

The relationship between the WRT and introspectionism is simple. A rather pedestrian consideration suggests that they are incompatible. If introspection lights the light of consciousness—i.e., if introspection lets one have one's qualia—then the qualia must be where introspection can access them. The innerness of introspection consists in the fact that its objects (not merely the introspective states themselves) are, in some sense, inside. Introspection is myopic—it's scope is limited to intracranial space. Externally located entities lie beyond its purview. The relocation$_w$ of qualia makes them into external, physical paint. All that remains in the introspectible domain are representations *sans* qualia. In introspecting these 'dequaliafied' representations one is not accessing qualia, *a fortiori* one is not accessing them in a way that makes it like something to do so.

Hence, relocation$_w$ precludes an introspective account of consciousness. Commenting on this point—a staple of the anti-introspectionist literature—Dretske has compared it illuminatingly, if unflatteringly, to the mistake of "inferring that because words (meaningful symbols) are in books, the meanings are also there." (Dretske 1995: 109) He concludes his reflections on books and brains by saying: "An internal scanner is as useful in mental affairs as would be a high-resolution camera in deciphering the meaning of coded text." (Dretske 1995: 109)

Note that this problem does not arise for all forms of introspectionism. Locke's version of introspectionism—the ur-version of introspectionism—does not suffer from this problem. Locke worked with unrelocated qualia. But it would be a mistake to think that that is what saved him. Views that combine relocation$_h$ with introspectionism are also immune to this particular problem—as is borne out by the theories of Rosenthal[115] and Pollock. The problem arises only if one attempts to combine the introspectionist idea with the relocation$_w$ of qualia. That this problem should arise only in these narrowly circumscribed conditions does not make it less serious. For these are precisely the conditions that define the position we are currently investigating.

How serious is the problem? Though I attach considerable weight to it, I doubt that there is no way out for the occupants of cell (3). I can see two escape routes. First, one might try to make the perceptual model of introspection more workable by adopting a more relaxed account of perception. Second, one might try to make introspection less myopic by, as it were, increasing the boundaries of inner space.

In previous chapters I meekly backed off whenever the argument led us right up to central questions in the philosophy of perception. In discussing the first strategy, I will follow this unedifying precedent and merely mention two versions of the perceptual model of introspection that appear to avoid the problem of myopia. The nutshell version of Armstrong's theory of perception says that perception is a matter of belief acquisition (see Sections 4.4.5 and 6.4.1). So if introspection is perceptual in nature then it too is a matter of belief acquisition. Since we are not limited to acquiring beliefs about what is in the head there should be no principled problem with extending the scope of introspection so as to encompass our environment. Relocated qualia—e.g., the redness on the tomato in front of me—are therefore fitting objects of the inner sense.[116]

Alston's theory of appearing (see page 119) affords another way out. If there is such a thing as being appeared to introspectively—which there probably isn't (see Section 4.4.4)—then all it takes for you to introspect an externally located quale is for the quale to appear to you in that way. Questions of inside and outside as well as questions about possible causal pathways fall by the wayside as

irrelevant. If the tomato's redness introspectively appears to you, then you introspect it.

These two approaches to perception are diametrically opposed to each other. I mention them here not to endorse them nor to suggest that Armstrong and Alston are respectively committed to these accounts of introspection. I use them merely to illustrate two possible ways in which the inner-sense theorist might attempt to bypass the problem I posed by enlarging the reach of the introspective faculty. Adopting the physical-paint scenario poses no problem to introspectionism if the inner sense can perceive the outer world.

The flavor of the second strategy is quite different though it comes to much the same in the end. What appeared to be a problematic consequence of the WRT—the problem of the inflation of consciousness (see Sections 7.3.1)—is put to use to solve the problem of myopia. If the mind is as big as the inflationary scenario would have it, then introspection—the intramental sense—has a virtually limitless reach. On this picture introspection is still limited to mental space. But with the release of mental space from the confines of the cranium this limitation no longer constrains. Thus the problem of myopia evaporates.

I have briefly canvassed some possible replies to the problem of myopia—the problem that arises if you ask the inner sense to access qualia located in the outer world. I find these replies unpromising. Thus, the attempt to combine relocation$_w$ with introspectionism strikes me as problematic. The space that the inner sense can represent and the space in which qualia are exemplified do not overlap. The inner eye cannot see physical paint. So it remains unclear how introspection will let us have our qualia.

I conclude from this discussion that the position marked out by cell (3) is significantly less promising than the introspectionism occupying cell (1). The strategy of combining wholeheartedly relocated qualia with an introspectionist account of how these qualia are had is unpromising.

8.5 Cell (4): External Qualia without HO Representation

We have finally arrived at cell (4) of Table 3 on page 184. This position—representationalism—combines physical paint (wholeheartedly relocated celluloid qualia) with a single layer of perceptual representation to tell the story about how we have our qualia. Fred Dretske and Michael Tye have presented the most fully developed versions of this sort of theory. My comments focus on Dretske's version of representationalism. I begin by reviewing some precursors of representationalism: naive realism and the position of G.E. Moore. I will argue that they do not so much explain as name a problem. Representationalism differs

from these views in employing the notion of a representation. But it appears that the introduction of this new term is not accompanied by a corresponding new insight. I close these comments on Dretske's representationalism by suggesting that Dretske's account may well explain the more humble phenomenon he set out to explain. Various things Dretske says about consciousness lead one to think that his notion of consciousness does not capture the phenomenon Nagel tried to capture with the 'what it's like' locution. If this diagnosis is correct, Dretske's representationalism should not be viewed as a theory of qualitative consciousness.

8.5.1 Precursors: Naive Realism and Moore

Before I turn to the specifics of Dretske's account, I want to spend some time with a precursor of contemporary representationalism. The view I have in mind is naive realism (in one of the many senses of this shifting term). According to naive realism, there is no room in the world for any mental paint in addition to the physical paint that adorns the objects in our environment. The rich phenomenology of which we become aware in visual perception is simply there, in front of us, unadulterated by any admixture of the perceiving subject. Such a naive realist theory of visual perception seems to yield a simple answer to the question how we manage to be conscious in the relocationist's world. To have a quale is simply to become directly aware of it. For it to be *like that* for me to experience the redness of the tomato is simply for me to become directly aware of the tomato's redness. In the case of vision, this sort of direct awareness is achieved by simply seeing the quale-bearing object.

There exists a battery of standard objections to naive realist theories of perception, concerned, for the most part, with problems about illusions (or non-veridical perception in general) and matters of perceptual relativity. These are genuine difficulties, and the countermeasures adopted by some of the naive realists are truly heroic. At this point it does not seem to me that either side can claim a definite victory. I do not mean to belittle these problems of naive realism by passing them over. But I want, instead, to focus on what is to my mind the most damaging objection to naive realism. My biggest complaint about the naive realist account is that I find it altogether unilluminating. Direct awareness functions as a primitive element in the naive realist account. And it is one of those primitives that do all the work the theory as a whole is supposed to do. I think it would be better to say that we are not dealing with a theory at all. What we have been given is a term that names a problem, not a theory that solves a problem.

To concretize this complaint a little bit we can recall Moore's account of sensation (see Section 7.3.1.1). Moore maintains that in a sensation two entirely

distinct aspects must be distinguished. First, there is the object—his example is *a blue,* where this is taken to be something "outside the circle of our own ideas and sensations." (Moore 1903: 27) And second, there is the element that is common to all sensations—consciousness, or awareness. In becoming directly aware of the blue we know it or, as I would put it, in thus becoming directly aware of the blue we get to have this quale.

So far so good. But now we want an answer to the hard question. How does this unique relation manage to do what it does? What about this relation makes it so that its holding between me and the blue results in its being like *that* for me to experience blue? How does the 'knowing' relation accomplish this amazing feat? Here we get the answer we usually get at this point, viz., no answer at all! The only explanation that Moore volunteers is this: "[My awareness] has to blue the simple and unique relation the existence of which alone justifies us in distinguishing knowledge of a thing from the thing known, indeed in distinguishing mind from matter." (Moore 1903: 26) As I understand this passage, it says that the knowledge relation is interestingly asymmetric. The object at the receiving end of the knowledge relation is always a mind, indeed, an object's capacity to stand at the receiving end of this relation is what makes this object a mind and distinguishes it from mere matter. And being known is the status conferred upon whatever object happens to be at the other end of the 'knowing' relation.

Note that this does not even begin to explain how the 'knowing' relation manages to execute its astonishing task of mediation. All we are being told is that our minds are the sort of thing capable of accomplishing such feats. Minds do it—that just about sums it up. Now, it is of course true that minds can do this sort of thing. But this isn't much of a theory. For it only states the fact that we (our minds) somehow accomplish the feat of having the blue, of enjoying a sensation *of* blue. This fact I grant, of course. But a statement of a puzzling datum is not an explanation of the puzzling datum. What we get here is not an explanation. It is merely the explanandum in disguise, dressed up to look like an explanans. This, then, is the sense in which I find naive realist accounts unilluminating. Explanations come to an end prematurely. The inquiry ends by mentioning a fact the existence of which is no less baffling than the phenomenon it allegedly explains.

It may seem that a discussion of naive realism and of Moore's analysis of sensations is quite beside the point in a discussion of representationalism. For neither the naive realists nor Moore help themselves to the central explanatory tool of representationalism, viz., the notion of a representation. This raises the question whether the appeal to representations substantially enriches the explanatory resources of the contemporary versions of naive realism.

We already have encountered a representation-based answer to the having-problem that is similar to representationalism in one important respect. Maloney's representational use theory of consciousness parallels representationalism in attempting to answer the having-problem without appeal to introspection. Both accounts claim that a single layer of representation allows us to have our qualia. The most characteristic feature of Maloney's theory—that it makes qualia into properties of the representations that reveal them—is the feature in which it differs most markedly from representationalism. The standard representationalist approach holds that representations do not bear the features they present to their user. Representations present us with qualia that are borne by objects other than the representations whose use allows us to *have* them. Relocationists$_w$, like Dretske and Tye, push the qualia all the way out onto the external object. So Maloney and the representationalists are at odds about the correct answer to the nature-problem. Maloney favors mental paint. The representationalists insist on physical paint. What we want to know is how relevant this difference is for possible solutions of the having-problem. We have seen that Maloney's theory does not manage to illuminate the darkness surrounding the having-problem. So the question before us is whether the switch to physical paint allows the representationalists to provide a better explanation of how qualia are had. I suspect that stripping the representations of the qualia they are supposed to present will not make this task any easier. We will still be faced with the question how tokening a representation of a quale—this time a representation that does not exemplify any qualia—makes it possible for me to *have* a quale as a result. How can the manipulation of inner symbols that are themselves devoid of phenomenal qualities bring it about that I thereby get to *have* the externally located phenomenal redness of the ripe tomato in the sort of way that leads to my ending up with a red-experience? Calling the process that works this miracle an act of representing does not really explain anything. It merely highlights the fact that the process is such that undergoing it makes the subject who undergoes it have the qualia that the relevant external object exemplifies. How such a feat is possible is a further question. And how well representationalism answers this question remains to be seen.

8.5.2 Dretske's Representationalism

In Section 4.4.4 we have seen Dretske's powerful critique of introspectionism. The negative part of his project—the critique of the idea that introspective representation is needed for, or capable of, igniting the light of consciousness—is admirable. But the positive part of the project is not similarly convincing. Those who can't wait to see the structure that Dretske is going to erect on the ruins of

introspectionism are in for a surprise. For he puts no new construction in the place previously occupied by introspection. He simply moves into the ruinous remains of introspectionism and declares victory. Dretske endorses the Armstrong/Lycan version of qualia. That is, perceptual qualia are properties of the intentional objects of these perceptions. In case the perception is veridical, the quale is identical with a physical property of the perceived object—the red quale of your tomato experience is the tomato's redness. Here is how Dretske puts the matter:

> Qualia are supposed to be the ways things seem or appear in the sense modality in question. So, for example, if a tomato looks red and round to S, then redness and roundness are the qualia of S's visual experience of the tomato. If this is so, then...if things ever *are* the way they seem, it follows that qualia, the properties that define what it is like to have that experience, are exactly the properties the object being perceived *has* when the perception is veridical. (Dretske 1995: 83-84)

So there are qualia and their very ordinary nature stands revealed. But one wants to know: How do we get from the tomato's redness to its being like *that* to see it? Why does it phenomenally seem like *that* to see a red tomato? Introspectionism tries to tell an explanatory story. Introspecting your tomato percept makes it like *that* to see a ripe tomato, introspectionism says. Dretske tells a story too. But its more of a just so story than an explanatory story. Experience, he tells us, represents systemically (represent$_s$). That is, an experiential state "derives its indicator function—and hence its representational status—from the system of which it is a state." (Dretske 1995: 12) With this terminology in hand, we are prepared for his answer to the question why it is like something to have experiences:

> Experiences are to be identified with states whose representational properties are *systemic*...As a result, experiences have their representational content fixed by the biological functions of the sensory systems of which they are states. How an experience represents$_s$ the world is fixed by the functions of the system of which it is a state. The quality of a sensory state—how things look, sound, and feel at the most basic (phenomenal) level—is thus determined phylogenetically. (Dretske 1995: 15)

No doubt phylogeny has something to do with it. We are the way we are because we became that way. Surely this is true. But the obviousness of this answer is merely a reflection of the cautiousness of this conjecture. What one wants to know is this: How did 'mother nature' pull it off? How does one put together a machine in such a way that it becomes like something for the machine to represent$_s$ certain features of its environment? This question is the interesting

one and it is still open. Why is it like *that* to see the ripe tomato? The answer that the represented properties are external rather than internal and that they are systemically represented rather than conventionally represented does not satisfy. This explanation leaves undiminished the bewilderment that originally prompted these questions.

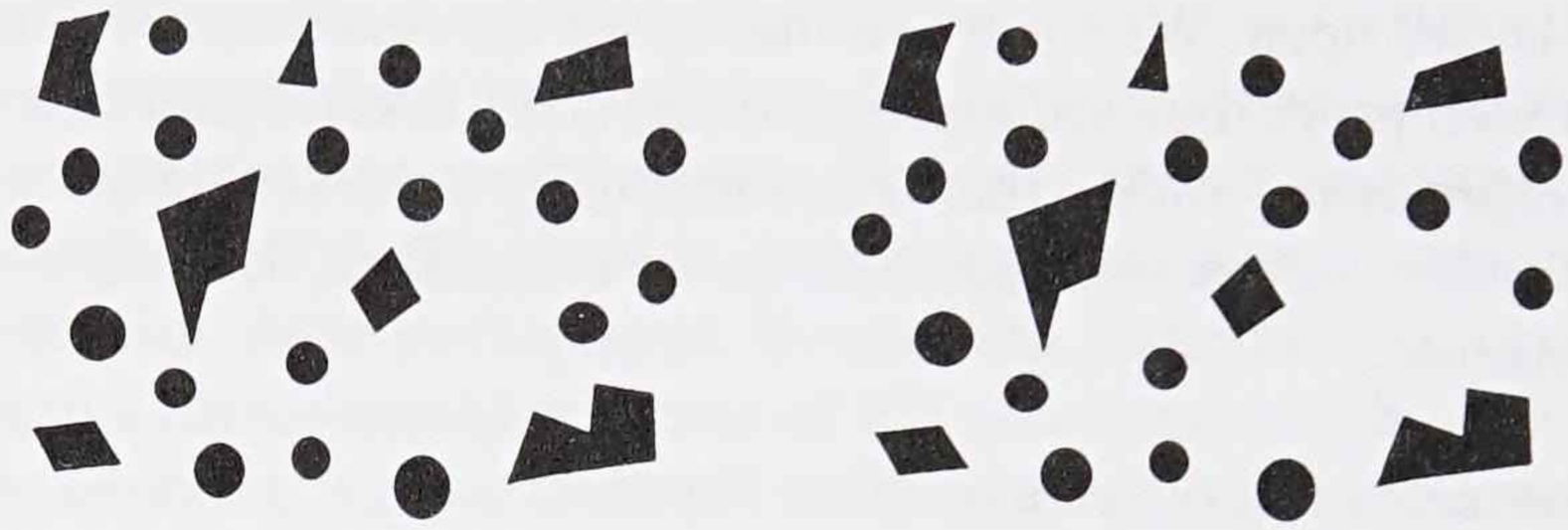
8.5.3 What Does Representationalism Explain?

Does Dretske perhaps see no deep further question here because he works with a less ambitious notion of qualitative consciousness? His many analogies with simple instruments and a superficial reading of the things he says about them—he talks about what it is like to be a speedometer, its qualia, experiences, and feelings—may suggest as much. But that would be too simple. Dretske surrounds all the relevant terms with quotes when talking about speedometers rather than people, thereby signaling that he does not want to be taken literally when he speaks in this way. In addition, he has stated an explicit criterion that determines which representations$_s$ are experiences: "They are those states whose function it is to supply information to a cognitive system for calibration and use in the control and regulation of behavior." (Dretske 1995: 19) This excludes speedometers, thermometers, and the like from the charmed circle. The view that Dretske does not talk about qualitative consciousness but about some watered down *ersatz* phenomenon seems difficult to maintain.

But there remains a lingering doubt that Dretske's notion of *conscious experience* fails to capture the phenomenon that Nagel's 'what it's like' locution points to. Most of Dretske's pronouncements on, and examples of, conscious experience create the impression that full-blown qualitative consciousness is his target. But the examples he uses to distinguish his view from introspectionism give one pause. In these examples it becomes clearest what is and isn't involved in Dretske's notion of a conscious state or a conscious experience. And after studying these cases carefully, one cannot help but suspect that what Dretske calls conscious experience is but a pale copy of the qualitative consciousness that makes it like something to be us.

I will state my worry by considering the case of Spot. (See Dretske (1993: 272-278)) Dretske provides the following viewing instructions for the experimental figure reprinted on the next page: "Glance at [the] figure long enough to assure yourself that you have seen all the elements composing constellation Alpha (on the left) and constellation Beta (on the right)…a second or two should suffice." (Dretske 1993: 273) Now turn the page and look at the figure.

If all went well Alpha and Beta looked the same to you. And that is so even though your conscious experience of Alpha differs from that of Beta. For Alpha



Alpha Beta

Figure 1. Dretske's spot

contains Spot—a spot that has no counterpart in Beta. And your experience of Alpha contained an experience of Spot—E(Spot)—that is not contained in your experience of Beta. Dretske presents E(Spot) as a conscious experience, as something that it was like something for you to have.

Well, was it like something for you to have E(Spot)? One's natural inclination is to say no. For had E(Spot) been like something for me, then I would not have judged Alpha and Beta to be the same. This reply may seem reasonable but Dretske will not have it. Whether there was something it was like for you to undergo E(Spot) is something you can determine only if you have an awareness of E(Spot). But the example was set up so that you would not have an awareness of E(Spot). Thus your failure to notice that it was like something to have E(Spot) does not show that it wasn't like anything for you to have E(Spot). For all you know, E(Spot) may have been a paradigm case of qualitative consciousness. And for all you know, E(Spot) may have been like nothing at all for you. One is tempted to think that E(Spot) is special or odd in this regard, that this quandary would not arise in ordinary cases of conscious experience. But that is a mistake. All of your qualitative consciousness is on a par with E(Spot). It so happens that you usually do have an awareness of your experiences. This misleads one into thinking that experiences are the same thing as experiences that we are aware of. but that, argues Dretske, is a mistake. The intrinsic character of an experience is prior to our awareness of it. Qualitative consciousness is 'in' your experience in a manner that is quite independent of any introspective activity on your part. Your experiences are qualitatively conscious but *you* may be completely in the dark as regards this dimension of theirs. The constant production of the 'technicolor phenomenology' of your experience is, in this respect, no different than the constant production of bile by your liver. As long as these processes go on unintrospected they take place outside of your awareness. They are nothing to you. E(Spot) does not come with the *for someone* effect that is characteristic of

qualitative consciousness. No directedness toward an apprehending mind is discernible here. This is a noteworthy result. It is completely at odds with the spirit in which the notion of qualitative consciousness was initially introduced. Surely this subterranean version of qualitative consciousness is not the striking and arresting feature of our experience that makes the mind-body problem hard. Once again our target has eluded us. The explanandum has quietly changed to something that appears to be far removed from the original baffling datum of consciousness.

Dretske's representationalism has much going for it. It is prefaced by the most powerful critique of introspectionism to date. By opting for physical paint it avoids the embarrassments of the colored-brain thesis. And by discarding the machinery of introspection it avoids the problem of myopia. If it fails to satisfy despite of all these virtues that is because it does not seem to offer a more illuminating answer to the having-problem than that provided by naive realism. The introduction of the notion of systemic representation does not dispel this impression. For one fails to grasp why making a quale the object of such a representation would eventuate in its being like something for the one who does so. But perhaps the real problem with Dretske's representationalism is located elsewhere. Some of Dretske's comments about qualitative consciousness make one think that the problem is not the explanans but the explanandum. If his target is a version of qualitative consciousness the enjoyment of which is like nothing for the subject, his account may well be a success. But this target is not our target. So we will have to look further.

8.6 Summary

This brings to an end the long march through the four cells of Table 3 on page 184. The four representation-based positions varied considerably. The details of the criticisms leveled against them varied to some extent. But the verdict with which I concluded the evaluation of each position varied not at all: "This account does not help us understand why having one's qualia in this way makes it like something to do so" was the monotonous refrain with which I ended every one of these discussions.

In stating the conclusion in this way I am probably guilty of making what Dennett has called the *first-person-plural presumption*. Described without pretense what I have done is really quite modest. I have laid before you the considerations that convince me that representation-based answers to the having-problem are unsatisfactory. I explained to you why I think that they do not ex-

plain qualitative consciousness. But I realize that my explanations are not proofs. So dissent is possible and likely.

How is one to proceed in a situation where all one has got is a subjective judgment of explanatory poverty? Even if we set aside the possibility that my judgment is grounded on some cognitive malfunction of mine, we still have no guarantee that my assessment will be shared intersubjectively. The pragmatic entanglements of the notion of explanation are notorious. And even if we all were to agree that representation-based answers to the having-problem are explanatorily barren, we could not infer that they are false or must be false. For true theories need not be explanatory and explanatory theories need not be true. (See Cummins (1983: 8-9)) What, then, am I to conclude from this discussion? Even in the best case scenario not much follows from my judgment of explanatory poverty. For all I have said it may still be the case that qualitative consciousness arises from the representation of relocated celluloid qualia. But if this is a truth it is one that I cannot comprehend. Perhaps there are incomprehensible truths one has to live with. But to acquiesce in such a 'solution' of a problem strikes me as something to be accepted only as a last resort. The present case does not call for such a desperate step.

Throughout the discussion in the last two chapters I tried to maintain an open mind about the relocation issue. In the previous chapter I tried to show that the qualophile has nothing to fear from a move to physical paint. More particularly, I argued that my earlier critique of introspectionism is not undermined by the relocation$_w$ of qualia. The discussion of the having-problem in the present chapter continued in the spirit of open-mindedness. I asked what it would take to have wholeheartedly or halfheartedly relocated qualia. But this led to an impasse. All respectable accounts of how relocated qualia are had are representation-based. But it turns out that I cannot comprehend such accounts. So if qualia are relocated, I cannot understand how they are had and my story about consciousness would have to end here. Open-mindedness about relocated qualia is what led to this impasse. So I face a choice: give up the attempt to understand consciousness or close the mind to relocated qualia. I propose to pursue that latter course. This sets the agenda for the next chapter. There I will try to show that there are good reasons to close one's mind to relocated qualia. If the argument of the next chapter is successful, we will only have learned a negative lesson: qualia are not relocated. This does not, of course, answer the question what qualia are or how they are had. These question I will take up in the last chapter.

8.7 My Position: A Preview

In Section 5.2.5 I made a first and, I fear, rather ill-fated attempt to sketch the outlines of the position I am working towards. Having just reviewed four rival accounts of the having-problem this may be a good place to undertake a second attempt to preview my own position. By positioning my view against the background of two of the representation-based accounts I hope to bring it into somewhat clearer focus.

My aim is to find a middle ground between two extreme positions. On the one side there is the cell (1) position (see Table 3, page 184). It combines mental paint with the machinery of higher-order representation. On the other side there is the cell (4) position. It favors physical paint and tries to make do without higher-order representation. Call the cell (1) position that combines internalized qualia with a bilevel representational structure 'thick internalism.' Call the cell (4) position that combines externalized qualia with a single layer of representation 'thin externalism.' Thick internalism relies on two factors to explain qualitative consciousness. There are the mental states that exemplify qualia, and there are the introspections of these qualitied states. The concerted working of these two factors results in its being like something to be you. Whenever one explains something by appeal to two factors there is a question of the division of explanatory labor. And thick internalism distributes the explanatory burden unequally. The quale of the first-order mental states is demoted into a humdrum physical property of a neural state. Nothing about these low-key, celluloid qualia foreshadows the spectacle of qualitative consciousness. The responsibility for that rests squarely on the shoulders of introspective representation. This special way of representing a quale lets full blown qualitative consciousness burst into the world, an astonishing special effect if ever there was one.

The thick internalist has a reason for believing that she can get away with her impoverished qualia. All the specialness she takes away from the side of the quale she adds back in on the side of introspective representation. The approach is based on the bet that the sophisticated machinery of introspective representation will cast more light onto the mystery of consciousness than the search for special intrinsic properties of qualia. So the recipe is this: keep the qualia simple and make the introspective representation of these simple qualia account for the specialness of qualitative consciousness. One may disagree with this approach, as Dretske and I do. But one has to admit that, in principle, this is a perfectly reasonable enterprise.

Thin externalism (as defended by Dretske) wants qualitative consciousness on the cheap. Dretske buys (and relocates$_w$) the 'qualia-light' of thick internalism but refuses to take the rest of the introspectionist machinery on board. The only

other notion he brings into play is that of systemic representation. But it remains unclear why representing, say, a red tomato systemically (rather than conventionally) results in its being like *that* to do so. Nor does the rider that only those systemic representations "whose function it is to supply information to a cognitive system for calibration and use in the control and regulation of behavior" (Dretske 1995: 19) help to advance our understanding of the origin of phenomenal consciousness. And Dretske enlists no other explanatory props with which to compensate for the explanatory poverty of the impoverished qualia. This exercise of unrestrained modesty leads to no good. A dilemma confronts the thin externalist. If the explanandum is genuine qualitative consciousness, then the explanatory means seem woefully inadequate—the explanatory gap looms ominously. And the toned down version of qualitative consciousness that the representational machinery can handle (exemplified by E(Spot)) is one that we are not interested in. So we face the common situation that a claim is either interesting and false or true and irrelevant.

This leaves me in a difficult position. I agree with the thin externalist objection against thick internalism that introspection plays no part in the production of qualitative consciousness. But I seem to be in agreement with thick internalism when I object to thin externalism on the grounds that what they serve up as qualitative consciousness is *ersatz* consciousness at best. What I want is an account that lies at a safe distance from both these extremes. The strategy must be to heed the valid criticisms that each side has leveled against the other while at the same time attempting to distill and preserve the correct insights buried in both. What the thick internalists have seen is that consciousness is always consciousness *for someone*, that consciousness has built into it a characteristic relation toward an apprehending mind. But the thin externalists have seen that the thick internalist way of cashing this idea—the fusion of consciousness and introspection—is misguided. In the process of distancing themselves from this mistake they overreacted and lost sight of the fact that consciousness is consciousness *for someone*. Thus we get an account of a curious sort of 'stealth-consciousness'—conscious experience that passes completely unregistered through the mind to which it belongs.

So how do you get the *for someone* effect without appealing to the machinery of introspection? My suggestion is to make the qualia do all the work. Celluloid qualia—these anemic 'qualia-light'—so eagerly embraced by thick internalists and thin externalists alike, cannot do the job. The qualia will have to be fattened up to the point where they no longer need rely on the help of introspective representations to become present to the mind of their owner. In other words, I want to vivify the qualia to the point where they become self-presenting. I want old fashioned silver-screen qualia. On this view, qualitative consciousness takes care

of itself. That is, one need not shine the light of introspection into the mind to bring forth the phenomenal landscape. Qualitative phenomena glow by themselves. That is what you get in return for making qualitative phenomena self-presenting.[117] On this view the existence and the feeling of a quale are the same and 'both' are prior to, and independent of, any introspective activity. Thus, the phenomenology associated with each of the 'ordinary' sensory modalities presents itself complete and unaided by any introspective support. Introspection is not needed to instantiate or to have of any of the visual, auditory, olfactory, gustatory, or tactile qualia. Nor, it would seem, are there any other kinds of qualia left over to guild and stain our inner perceptions. This rests all of the explanatory burden on the beefed-up notion of a quale. The bet that underlies the approach is very nearly the opposite of the introspectionist wager. My recipe is this: forget the machinery of representation and make the intrinsic structure of the quale account for the specialness of qualitative consciousness.[118]

Here is a very simple minded analogy. If you take a late night drive down the strip, you encounter two basic kinds of signs: those made of painted wood illuminated by a spotlight and neon signs that shine by themselves. There are also unilluminated wooden signs; but those you do not see. I will use these three kinds of signs to bring out the contrast between thick internalism and thin externalism and to suggest the sort of view that I want to defend. The thick internalists's strip is made up of only wooden signs with spotlights directed upon them. This struck the thin externalists as rather old fashioned. They noticed that spotlights are no good—they are expensive, they break, they illuminate the wrong places, etc. Moreover, they pointed out that the signs were brightly painted in an eye-catching way. So they had all the spotlights removed from the thick internalists's strip, reasoning that the cruising experience would be none the worse for that. The thin externalists's strip turned out to be a rather dull affair populated only by a few shady customers. I share the thin externalists's antipathy to spotlights. But I find the thin externalists's strip depressing. I therefore propose to revitalize the strip by tearing down all the wooden signs and by replacing them with neon signs that shine by themselves, unaided by spotlights. Perhaps, then, I should call my view 'the neon theory of consciousness.'

CHAPTER NINE

Denying Relocated Qualia

In Chapter Seven (especially Section 7.4) I argued that the case for the relocation theses consists of two parts—one positive, the other negative. We have already seen the positive part. It consists of the second appeal to phenomenology (Section 7.4.1) and the claim that there is no Mental home for qualia (Section 7.4.3). In this chapter I discuss the negative part. To show that the relocation of qualia is feasible one must show that the relocation theses can be defended against the argument from illusion and the argument from atomism.

In Chapter Seven I argued that both versions of the relocation thesis are supported by the consideration that there are no Mental qualia-bearers and that the second appeal to phenomenology does not favor the WRT over the HRT, notwithstanding the fact that it is designed to do so. So it seemed that the WRT and the HRT were equally well supported by the positive arguments in their favor. In this chapter I will argue that the negative part of the case for the relocation theses preserves the evidential symmetry between them by proving them both untenable.

The first of the two anti-relocationist arguments presented here—the argument from illusion—may seem to break the symmetry between the WRT and the HRT by strongly favoring the HRT. The argument focuses on nonveridical experience. The sort of question I will ask is of the following form: Which object bears the red quale of your tomato hallucination? This embarrasses the relocationist$_w$, for there appears to be no suitable external object to do the job. But nonveridical experience poses no problem for the relocationist$_h$. Hallucinating involves mental states that are no less real than those involved in veridical experience. And these inner, mental states bear the experienced qualia. I end this part of the chapter with the claim that, contrary to first appearances, the relocation$_h$ of qualia is no better off than its wholehearted twin.

The second argument—the argument from atomism—cuts equally against both relocation theses. I argue that no part of the physical world—be it outside or inside the head—is fit to bear qualia. This means that the relocationists$_w$ are in trouble even in the domain of veridical experience. It also means that the mental qualia-bearers of the relocationists$_h$ are problematic. For these mental states are

physical states. The external physical tomato is as unfit to exemplify the red quale of my veridical tomato experience as is the mental state with which I represent it. I conclude that both relocation theses have to be given up.

9.1 The Argument from Illusion and Nonexistent Qualia-Bearers

Illusions, hallucinations, dreams, and mental images pose a problem for the wholehearted relocation thesis (WRT).[119] A person going through any one of these non-veridical experiences enjoys phenomenal qualities, yet there is no corresponding physical object to which these qualities might be ascribed. There is no rat in front of the person who hallucinates a pink rat. Thus, it is a mistake to locate the phenomenal pinkness on the rat; there simply is no such rat, pink or otherwise.

Many arguments have been built around these various kinds of nonveridical experience. Collectively they go by the name 'the argument from illusion.' This argument is a staple of the anti-materialist literature. Supposedly it proves the existence of sense-data. Materialism cannot accommodate sense-data. Therefore the argument from illusion is taken to refute materialism. Be that as it may, most everybody seems to agree that the kinds of phenomena to which the argument from illusion draws our attention—illusions, hallucinations, dreams, and imaginations—pose at least a *prima facie* problem for any theory that would locate phenomenal properties on the external physical object. Where there is no external physical object, the attempt to move the phenomenal properties out of the mind and onto the external object is doomed. The charge against the relocationist$_w$ comes to this: in the cases of non-veridical experience, he cannot provide a home for qualia.

9.1.1 The Route to Nonexistent Qualia-Bearers

One reply, which has found a number of defenders in the recent literature, consists in the startling claim that even in those instances in which the ostensible object of perception does not exist, the experienced phenomenal qualities are borne by a physical object, albeit one that happens not to exist. According to this view, the division of qualia-bearers into real physical objects and real nonphysical sense-data does not exhaust all possibilities. The mistaken acceptance of this dichotomy is rooted in the provincial assumption that the actual physical objects are all the physical objects we can work with. But once we rid ourselves of this limiting prejudice, it is simply not true that there is a shortage of suitable physical objects to bear the qualia we experience in illusions, hallucinations, imagings, and after-images. On top of all the actual physical objects we now have all the

nonactual physical objects to bear our qualia. Thus, the materialistically correct answer to the question which physical object bears the red quale of my tomato hallucination is this: the object bearing my hallucinatory red quale is a tomato, physical through and through; as it happens, it is a nonactual tomato, but it is none the worse for that. The upshot of this maneuver is that qualia of illusions, hallucinations, dreams, etc. no longer pose a threat to the WRT. For every experience—be it veridical or nonveridical—the relocationist$_w$ can supply an external physical object—actual or nonactual, as the case may be—to bear the qualia of the experience.

The defense of this position proceeds by emphasizing the intentionality of sensation.[120] A sensation is always a sensation *of* something, regardless of whether the sensation's ostensible external physical object does or does not exist. If you hallucinate a pink rat, there is no pink rat in front of you; nevertheless, your sensation is a sensation of a pink rat. Just because there is no rat in front of you, your sensation does not turn into a sensation of nothing, or into a sensation of a table or of a wall, or of whatever else might really be in front of you. Your sensation remains the sensation of a pink rat whatever else might be true of your immediate environment. In other words, your sensation has a pink rat as its intentional object, which may or may not exist. In this manner the intentionality of sensation comes to the aid of the relocationist$_w$. No matter whether we are dealing with veridical or non-veridical experience, every sensation has an object, albeit a merely intentional object in the case of non-veridical experience. And so long as there is an object of any kind there is a place for the evicted phenomenal qualities. Thus, non-veridical experience does not embarrass the relocationist$_w$ into admitting homeless qualities into his ontology: the intentional object is the proper home of qualia.

Though this position may strike one as peculiar, it has been endorsed by a number of recent writers. In the course of discussing the question whether hallucinations force us to acknowledge the existence of mental paint, Harman draws our attention to the fact that "the object of awareness is an intentional object that need not exist." (Harman 1989: 841) Concerning a situation in which Alice is hallucinating a ripe tomato, he writes:

> In the case of hallucination Alice is aware of something red, but the red thing is a nonexistent physical thing rather than an existent mental thing...what Alice sees or seems to see when she hallucinates is the redness of a tomato. Maybe there is no redness in the world before her, but that does not mean she sees mental redness. It only means that she is seeing (or seeming to see) something that does not exist. (Harman 1989: 841)

Armstrong proposes a similar solution to the problems posed by non-veridical experience. He recommends that the causal theorist who is confronted with the argument from illusion should proceed as follows:

> He should deny that when something physical looks green to somebody, but is not green, or where somebody images something green, then the sensory quality of greenness is present. The causal theorist can admit that there is a sense in which sensory illusion, or the having of such images, involves something green. But (a) the something is an ordinary physical something; and (b) it is a merely *intentional,* not a real, object. (Armstrong 1984: 171)

Once the role of the intentional object is understood, "there is no call to treat illusory or imaged sensible qualities, and in particular color, sound, taste, smell, heat or cold, as actual qualities of actual entities." Inasmuch as we are concerned with actual objects and their properties we can therefore say that "the illusory qualities do not qualify anything at all." (Armstrong 1984: 174)

Lycan presents an interesting variant of this idea. It is clear that he, too, defends the view that nonexistent objects are a fit home for qualia. However, he creates the impression that the relevant nonexistent objects are quite different in nature than those that Harman and Armstrong contemplated. Where they speak of nonexistent tomatoes and vines, etc., he speaks of nonexistent phenomenal objects. Talk of phenomenal objects is traditionally associated with talk of sense-data and their ilk. And the attempt to avoid sense-data is one of the main motivations of the WRT. Lycan's talk of phenomenal objects therefore suggests that his appeal to merely intentional objects does not belong with the WRT (or any form of relocationism, for that matter). But this contrast is more apparent than real. Lycan is not interested in enlisting nonexistent nonphysical entities as qualia-bearers. Like Armstrong and Harman, he recruits nonexistent physical objects as the intentional, qualia-bearing objects of nonveridical experiences. All these objects share with traditional phenomenal objects is their playing a certain role, viz., that of exemplifying qualia. So we must not let ourselves be misled by Lycan's language. He, too, defends the view that intentional objects of sensations—construed as nonexistent physical objects—save the WRT from the problems posed by nonveridical experience. Here is how Lycan states his case. He considers a person (with the rare and beautiful name Leopold) who is suffering a green afterimage:

> Leopold's green patch is not here in this room with Leopold, because it is not in our world, the actual world, at all. I take the view...that phenomenal individuals such as sense-data are intentional inexistents à la Brentano and Meinong. It is, after all, no surprise to be told that mental states have intentional objects that may not exist. So why should we not suppose that afterimages and other sense-data are intentional objects that do not exist? If they

> do not exist, then—*voila!*—they do not exist; there are in reality no such things. And that is why we can consistently admit that phenomenal-color properties qualify individuals without granting that there exist individuals that are the bearers of phenomenal-color properties. (Lycan 1987: 88)

Note that so far Lycan has not explicitly claimed that phenomenal objects are nonexistent *physical* objects. But Lycan's elaboration of his position makes it amply clear that these reconstructed sense-data are regular, if nonexistent, physical objects. It is much to Lycan's credit that he sees and addresses the problems this causes. (See Lycan (1987: especially 91-93)). Physical objects in other possible worlds inherit whatever problems physical objects in this world may confront when called upon to exemplify qualitative features like color. As I will argue in Section 9.3, these problems are considerable. Pushing qualia-bearers into a nearby possible world without changing their nature does not alleviate these problems in the least.

This last observation raises an interesting question: Must a materialist insist that the otherworldly qualia-bearers are physical? A modest materialist position could acknowledge the fact that nonexistent nonphysical objects 'enter into' some of our mental states. This would leave untouched the central claim of materialism that all objects of this world are physical objects. To such a materialism nonphysical objects would be acceptable as long as they do not exist. Though such a position satisfies the letter of the materialist doctrine, it may strike one as offending against the spirit of the creed. Nevertheless, it should not be discarded lightly for it has one great advantage (even if it seems to be gained by theft rather than honest toil). By keeping the (merely) intentional object nonphysical, it avoids all the problems that arise from the fact that physical objects do not seem to agree with qualia. Nonphysical objects and qualia, on the other hand, seem to be made for each other. Of course, this is so because they *were* made for each other. Or rather, nonphysical objects, like sense-data, were posited so as to play the role of qualia-bearers. It is therefore no wonder that the question how they can manage to bear qualia is never asked. (It is less clear whether this question should never be asked). To repeat—this is not the position Lycan wishes to defend. It is a position in logical space and unlike many such positions it has some features that recommend it.

There is, then, a considerable consensus about how the WRT can best deal with the problems raised by nonveridical experience. The key insight is that all sensation is intentional and that intentional objects need not exist to fulfill their function. These nonexistent intentional objects, understood as physical objects, are the natural loci of the qualia we have in nonveridical experiences. This is the view that I am going to critique throughout the first part of this chapter. But before moving into this critique I want to set aside a deflationary account of inten-

tional objects. Contrasting this deflationary account with the one I find in the authors discussed above, may help to get a somewhat firmer grip on the elusive notion of an intentional object.

9.1.2 Intentional Objects—Mere Manners of Speech?

Throughout, I will take intentional objects to be objects in their own right. They might be physical or nonphysical, actual or nonactual—but no matter what their nature is, I will always assume that they are objects of a subject's mental act and that they are capable of exemplifying qualia.

Other authors use the expression 'intentional object' in a much more non-committal way. They deny that an intentional object is an object in its own right. According to this view, intentional-object talk is merely a manner of speaking. This sort of talk does not commit one to holding that there are objects of any kind to which the subject relates itself through an act. Here is how J.L. Mackie formulates this view: "To say that there is an intentional object of a certain sort is only to say, in what could be a misleading style, that that is how things look (or feel, or sound, and so on) to the person in question." (Mackie 1976: 48) If, for example, we consider a person, Tom, who feels a pain, sees a color, or has some other sort of experience, we should describe the situation by saying that "the only entity involved is…Tom's having an experience of a certain sort. Talk about its intentional object can be no more than a way of characterizing it, of saying what sort of an experience it is by indicating its content." (Mackie 1977: 110-111) We are led to mistake the content of the experience for an object in its own right and to endorse a relational theory of experience because

> …this content, though it is really only a feature of the experience, though it really only makes it the particular sort of experience it is, presents itself as if it were a more or less distinct object to which the subject is related, this whole relational situation then being the experience. (Mackie 1977: 111)

In short: contents masquerade as separable objects (Mackie's phrase). Superficial thinkers fall for the masquerade and treat intentional objects as genuine objects. The more enlightened philosophers see through the masquerade and employ the expression 'intentional object' only as a manner of speaking, thereby canceling all commitments to there being such objects (whatever their precise nature may be).

My disagreement with Mackie is not substantial. In Chapter Ten I present my own reasons for adopting a nonrelational analysis of experience. But in the present discussion I want to be understood literally when I speak of intentional objects. That is, in the present discussion I take intentional objects to be *bona fide* objects, not mere manners of speech. The point of this discussion is to sort

through a number of proposed accounts of intentional objects and to evaluate them with regard to their suitability as qualia-bearers.

The literal reading of the expression 'intentional object' is well motivated, for it allows one to see distinctions that are obscured by Mackie's way of talking. Consider, for example, the contrast between the sense-datum theory and adverbialism. Sense-data are intentional objects par excellence. They are the qualia-bearing objects to which the perceiving subject is directly related. They are also highly problematic. The big lure of adverbialism is its claim to get by without postulating these problematic intentional objects. There is, then, a sense in which the sense-datum theory does, and adverbialism does not, posit intentional objects. This is the sense in which I want to be understood when talking about intentional objects. Mackie's approach fails to capture this important sense of the expression 'intentional object.' According to him, experiences have intentional objects no matter how they are construed. For Mackie, there is in this respect no difference between a sense-datum account of experience and an adverbial account of experience. On both accounts experiences are of a certain sort, i.e., they have intentional objects.

A second, and much more practical, reason for making this choice is the fact that the philosophers whose work I discuss take intentional-object talk seriously. To engage them on their own ground, I must follow their example. Consequently the following discussion of intentional objects relies on the literal or robust reading of the expression 'intentional object.'

9.2 Problems With Intentional Objects

I will now argue that intentional objects are a poor home for qualia; especially so, if they are introduced in an attempt to save the WRT. Not just any sort of intentional object will be able to do the work the relocationist$_w$ needs it to do. It will, therefore, be useful to remind ourselves of the kinds of constraints that intentional objects must satisfy in order to be of any use to the relocation$_w$ theorist. First, intentional objects must be nonmental. To push the qualia out of the mind—not merely out of the Mind—is the goal of the relocation$_w$ move. So the relocationist$_w$ must insist on intentional objects that are physical and nonmental. Second, intentional objects must be of a kind that can plausibly be said to exemplify phenomenal properties. Unextended point-particles, for example, are the wrong kind of bearers for phenomenal qualities like redness. And third, the intentional objects must be such that their exemplifying phenomenal properties can contribute to the explanation of our conscious experience. If, for example, we were to push them into some world to which the experiencing subject bore no

relevant relation, then something's being red in that faraway world could not explain how this quality affects the subject's experience.

The first constraint on suitable intentional objects is obvious. But the last two merit further comment. The relocationist$_w$ relies on intentional objects to overcome the problems posed by nonveridical experience. This will work, only if the intentional objects, as he conceives of them, are capable of providing a home for phenomenal qualities in cases in which there are no external physical objects to bear them. This puts a constraint on possible candidates for intentional objects: they must be objects of a kind that can exemplify phenomenal properties. Thus, a question we will have to pose to any theory of intentional objects will be this: Are the objects posited by the theory of a kind that can exemplify phenomenal qualities? I will call this the 'exemplification-problem.' The exemplification-problem is, of course, just one version of the familiar nature-problem. The move to intentional objects does not prevent this problem from arising. Every theory of intentional objects to which the relocationist$_w$ helps herself will have to deal with the nature-problem.

Solving the nature-problem for intentional objects is only one half of the relocationist's$_w$ intentional-object theory of nonveridical experience. The relocationist$_w$ must also give an account of the relation obtaining between the qualitied intentional object and the experiencing person. That person experiences the particular phenomenal qualities she does in virtue of her relation to the qualia-bearing intentional object. To the extent that the relocation$_w$ theorist is naturalistically inclined, he will want to spell out this relation in causal terms. Thus we arrive at a second constraint on possible candidates for intentional objects: they must be objects of a kind that can interact causally with experiencers to produce the requisite experiences. Hence, a further question we will have to pose to any theory of intentional objects will be this: Are the objects posited by the theory of a kind that can enter into the relevant causal relations with experiencers? I will call this the 'causation-problem.' The causation-problem is, of course, merely a component of the having-problem. Given a naturalistic framework, a solution to the causation-problem is a necessary, but not sufficient, step towards a solution of the having-problem.

The nature-problem is logically prior to the causation-problem. If you fail to solve the first, you fail completely and the causation-problem does not even pose itself. The following review of various construals of intentional objects will show that they do not satisfy the three constraints on qualia-bearing intentional objects enumerated above. Such an argument by elimination falls short of a proof that intentional objects of the kind required by the WRT cannot are unavailable. But it does show that the attempt to save the WRT by an appeal to intentional objects is an extremely problematic undertaking.

9.2.1 Intentional Objects à la (Early) Brentano

The point of the WRT is to transfer the phenomenal properties out of the mind and onto the mind's nonmental object. But transferring the phenomenal properties onto the intentional object does not achieve this goal. For, according to the traditional conception of intentional objects, they are mental objects. So the whole maneuver is pointless as far as the WRT is concerned. The current interest in intentional objects dates back to Brentano's famous characterization of mental phenomena:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgment something is affirmed or denied, in love loved, in hate hated, in desire desired and so on. This intentional in-existence is characteristic exclusively of mental phenomena. No physical phenomenon exhibits anything like it. We can, therefore, define mental phenomena by saying that they are those phenomena which contain an object intentionally within themselves. (Brentano 1973: 88-89)

In this well known passage—albeit "not well known for its clarity," as Tim Crane (1995: 31) points out—Brentano does two things: he presents a thesis about the distinguishing characteristic of the mental; and he presents an ontological thesis concerning the kind of existence that belongs to intentional objects. It is the latter thesis with which we are concerned. What does Brentano's ontological thesis tell us about intentional objects? Here is how R.M. Chisholm, the foremost contemporary reviver of Brentano's doctrine of intentionality, summarizes what the ontological doctrine of intentional inexistence tells us about the intentional object of a man who thinks about a unicorn: "It tells us, first, that the object of the man's thought *is* a unicorn. It tells us, second, that this unicorn is not an actual unicorn (for there are no actual unicorns). And it tells us, third, that this unicorn has a certain mode of being other than actuality." (Chisholm 1970: 136) The first two points pose no threat to the WRT. If the relocationist$_w$ is to locate a new home for qualia he, too, must insist that the man is related to a unicorn that is not actual. It is the third claim that proves to be problematic if spelled out further. For it is here that the mentality of intentional objects comes into play. Chisholm's interpretation of the particular mode of being that is proprietary to intentional objects makes that abundantly clear:

> Whatever has this mode of being—called "intentional inexistence" or "immanent objectivity"—is an entity that is mind-dependent and therefore appropriately called an *ens rationis,* in the traditional sense of this term. The intentionally inexistent unicorn is an entity that is *produced* by the mind or intellect; it comes into being as soon as the man starts to think about a unicorn and it ceases to be as soon as he stops. (Chisholm 1970: 136)

Brentano provides the WRT with an intentional object, but it is an object of the wrong kind. Since it is a mental object, the relocationist$_w$ cannot accept it as a suitable support for phenomenal qualities.[121]

9.2.2 Intentional Objects à la Meinong

The better part of the ensuing discussion about intentional objects was driven by an attempt to rid oneself of the troublesome ontological liability of mental inexistents. Brentano himself was among the first to attempt such a purge. But the relocation$_w$ theorist should be wary of this laudable drive towards ontological parsimony. For he must make sure that he does not end up empty handed. That is, he runs the risk of being left with nothing that could plausibly serve as support for qualia once intentional objects have been eliminated or reduced. Suppose, by way of an example, you established that our intentional attitudes are really directed towards linguistic entities. What is the relocation$_w$ theorist to do under these circumstances? He cannot plausibly maintain that the phenomenal qualities of experience are intrinsic properties of linguistic entities. So he has to resist this form of linguistic reduction of intentional objects. Only those accounts of intentional objects will merit the relocationist's$_w$ attention that construe intentional objects as nonmental objects to which phenomenal properties can plausibly be attributed. I will now argue that the attempt to construe intentional objects as Meinongian objects—objects beyond being and nonbeing—succumbs to just this problem. The objects which this account furnishes cannot do the kind of work that the WRT requires an intentional object to do.

One way to guarantee that you are not dealing with a mental object is to make sure that you are dealing with an object that does not exist: if it does not exist, it does not exist in the mind. There are various attempts to make sense of nonexistent objects. One is due to Alexius von Meinong, Brentano's most famous pupil. Lycan gestures towards a Meinongian account of intentional objects. But I doubt that the relocationist$_w$ would be well advised in following up this lead. The objects Meinong singles out as the intentional objects of sensations are remarkably unfit for the job the relocationist$_w$ needs them to do. According to Meinong, the objects of our sensations (*Empfindungsgegenstände*) are "homeless objects" (Meinong 1907: 8-14); homeless "because there would seem to be no place for [them], either in Plato's heaven or on earth." (Chisholm 1973: 37) They are

among Meinong's famous *außerseiend* objects—objects that have no form of existence, being, or reality whatever—of which it is, nevertheless, true that they have certain properties. Their *Sosein* is independent of their *Sein*. Can these objects that are, in another one of Meinong's phrases, beyond being and nonbeing, be considered a satisfactory home for the extradited phenomenal properties?

Here I can do no more than admit my inability to understand how homeless objects are supposed to afford a home for qualia. I can best express my puzzlement by asking H.H. Price's question: "How can an unreal tomato be really red?" (Price 1954: 105) On the Meinongian construal of intentional objects, both, the nature-problem and the causation-problem, cry out for answers. The Meinongian relocationist$_w$ has to tell us what it is for a homeless object to be red, and what good a red homeless object would do us.

One may be tempted to answer the first question in the language of possible worlds: "For a homeless tomato to be red is for there to be a red tomato in some possible world that is not the actual world. And what it is for a tomato to be red in a nonactual possible world is no more clear or obscure than what it is for a tomato to be red in the possible world that is the actual world." But I think no answer along these lines will do. For a tomato that exists in some possible world is not a homeless tomato; its home is the possible world in which it exists.

This rejoinder may prompt a second attempt to answer the first question in the language of possible worlds: "A homeless red tomato is one that does not exist in any possible world." But this will not do either. First, it seems to me that this answer does not advance our understanding of the question with which we began. It is no more clear what it is for a tomato that does not exist in any possible world to be red than what it is for a homeless tomato to be red. Whoever has a problem with homeless red tomatoes will also have a problem with the redness of tomatoes that do not exist in any possible world. Second, the answer that we are now considering has turned our red tomato into an impossible object, which is more than we had bargained for. Meinong would, of course, agree that there are impossible homeless objects—he makes much of the round square—but he would deny that all homeless objects are impossible objects. Therefore, the notion of a homeless object cannot be cashed out in terms of the notion of objects that do not exist in any possible world. For all such objects are impossible objects.

But let us assume that there is a good answer to be had to the first question and that it is an unproblematic affair for a homeless object to be red. Can the relocation$_w$ theorist rest content with the claim that nonveridical experience acquaints us with phenomenal properties of homeless objects? Does the Meinongian have an answer to the causation-problem? What, short of a miracle, could make it the case that I experience phenomenal red *in virtue of* there being a

red homeless object, an object for which there is no place "either in Plato's heaven or on earth"? No naturalistically respectable relation, it seems, could bring me and a homeless object together in such a way that I get to have the qualia it instantiates. The principle underlying this worry has been lucidly expressed by McGinn: "CAUSAL RELATIONS ARE ONLY AS NATURAL AS THE TERMS THEY RELATE. The notion of causality is not, as it were, natural in itself: *it* will be spooky in proportion as its terms are. Occult entities breed (and need) occult causation." (McGinn 1991: 52) According to McGinn, it is the absence of any conceivable mechanism that makes these sorts of causal relations problematic. This consideration applies in full force to the proposal to construe intentional objects as homeless objects. Nothing short of a Moorean 'the mind just does it' theory (see Section 8.5.1) 'sheds light' on how homeless qualia are had. And the Moorean theory itself is no better off in this respect—it names a mystery without unravelling it.

9.2.3 Intentional Objects and Possible Worlds

I have argued that the framework of possible worlds does not offer adequate interpretative resources for Meinongian intentional objects. However, from this it does not follow that the apparatus of possible worlds can shed no light on intentional objects. The connection between these two domains will appear more clearly if we proceed as follows: instead of using the apparatus of possible worlds to interpret the notion of an intentional object arrived at independently, we will try to fashion our notion of an intentional object along the lines suggested by the framework of possible worlds. On this approach it is natural to take a merely intentional object, like the hallucinated red tomato, to be an object that fails to exist in the actual world but does exist in some other possible world—a non-actualized possible or a mere possibile. Since we are now working without a preconceived view about the ontological status of intentional objects, we will stipulate that intentional objects exist in whatever way objects in non-actualized possible worlds exist. But this claim is too unspecific as it stands. For there is no *one* way in which mere possibilia exist. Different accounts of possible worlds yield different stories about mere possibilia. And since the differences between rival accounts of possible worlds are profound, the resulting accounts of mere possibilia differ significantly too. Using possible worlds to explicate the notion of an intentional object does not yield an unequivocal result. That is, whether the project of explicating the notion of an intentional object by relying on the framework of possible worlds can be judged a success will depend, to a large measure, on which particular account of possible worlds you choose. There are numerous extant accounts of possible worlds. Following

David Lewis, I will distinguish only two kinds of stances towards possible worlds: the Lewisian realist stance and the ersatzist stance. Depending on whether you are a Lewisian realist or an Ersatzist your account of the nature of mere possibilia will differ greatly. I will argue that in the end it does not matter which branch you choose. Neither one of the resulting notions can do what the WRT demands from its intentional objects. But they fail for different reasons. I will argue that theories of intentional objects conceived in the spirit of Lewisian realism can solve the nature-problem but fail to solve the causation-problem. The ersatzist theories of intentional objects either fail to solve the nature-problem, and therefore don't even get a shot at solving the causation-problem, or, if they do provide an answer to the nature-problem, they cannot deal with the causation-problem.

9.2.3.1 Lewisian Realism

For the Lewisian realist all possible worlds exist in the most robust sense of the word. Possible worlds are and contain physical objects just like the ones we encounter every day. Hence mere possibilia are regular physical objects and can exemplify phenomenal properties with the same ease with which actual objects exemplify them.[122] Since, according to this account, mere possibilia are ordinary physical objects, they also would seem to be the right kind of objects for us to be causally related to in experience. But here the Lewisian realist encounters a hitch. Possible worlds and the possible objects they contain are spatio-temporally and causally separated from the actual world. No physical transaction of any kind transpires between possible worlds. If intentional objects were Lewisian possibilia, then suffering a nonveridical experience would not be like anything for the experiencer. For there is no naturalistically respectable way for her qualia-laden intentional objects to color her experience. All the many worlds full of brightly colored possibilia are in vain. Becoming a Lewisian realist about intentional objects does not advance the cause of the WRT.

9.2.3.2 Ersatzism

The case of the ersatzers is less straightforward, if only because their numbers and their disagreements are great. I will therefore not focus on individual ersatzers but on a general feature shared by many of their constructions. I realize that this procedure has the drawback of appearing to be an attack on a straw man specifically constructed so as to be easily overcome.

The common trait of ersatz worlds is that they are understood to be abstract objects, properties, propositions, or states of affairs, for example. If we focus on states of affairs we can say, following Alvin Plantinga, that a possible world is a maximal possible state of affairs. That is to say, a world w is a state of af-

fairs such that for every state of affairs S, *w* either includes S (S must obtain if *w* obtains) or precludes S (S cannot obtain if *w* obtains). (See Plantinga (1976: 258)) What does this setup have to offer the relocationist$_w$ in his effort to secure an intentional object for the person hallucinating a pink rat?

One might try the following: "Our world does not include the state of affairs of *the rat's being pink*. But the state of affairs of *the rat's being pink* is included in some other possible world. This latter state of affairs is the intentional object of the person who hallucinates a pink rat."

This will not do, however. Qua abstract objects states of affairs are the wrong kinds of objects to stand in for intentional objects. What would it be for an abstract object like a state of affairs to be phenomenally pink? This is not to deny that the state of affairs of *the rat's being pink* obtains. It is just to deny that this state of affairs itself or its obtaining could be phenomenally pink. *The rat's being pink* is not pink, nor is the obtaining of *the rat's being pink* pink. What one wants to say is this: if anything is pink it is not the state of affairs itself but rather the rat which, so to speak, is a constituent of the state of affairs that consists in the *rat's being pink*. But which rat is that? It is not one of the rats that exist in the actual world: there is no rat, pink or otherwise, in the hallucinating person's environment. Should we say, then, that it is quite literally an otherworldly rat; one that does not exist in this world but does exist in some other possible world? Now, if this otherworldly rat is to be the bearer of the phenomenal pinkness which our hallucinating person experiences, then it will have to be a concrete rat of the sort that David Lewis believes in. But we have already seen why Lewis's concrete possibilia cannot serve as intentional objects. In an attempt to retreat from the concrete possibile, one might try to 'thin out' the problematic concrete rat into something like the state of affairs of *the rat's being*. But now you are back at abstract states of affairs that will not support pink (or any other) qualia. 'Thinning out' the otherworldly rat even more until one ends up with one of Meinong's homeless rats will not improve matters. I have already argued that homeless objects cannot serve as intentional objects. Thus, I do not see how embracing ersatzism will help the relocation$_w$ theorist. None of the states of affairs containing rats, nor any of the types of rats that might go into the making of such states of affairs seemed to be of the kind that could comfortable exemplify phenomenal pinkness. It would seem that the ersatzist cannot solve the nature-problem.

The ersatzist will impatiently reply that I have overlooked the most straightforward and obvious solution of the nature-problem in the framework of ersatzism. What one should say is simply this: the rat in question is one of the actual world's rats. Had things been relevantly different there would have been an actual pink rat in our hallucinator's field of vision. That, then, is the rat that bears

the phenomenal pinkness that the hallucinator experiences. Since this rat is a respectable physical object, its being pink is no more problematic than any physical object's being pink. Thus the nature-problem is easily dealt with.

But if this is the route the ersatzist wants to take he gets himself impaled on the causation-problem. How can my actually experiencing a pink quale be explained by the fact that a rat which, as a matter of fact, is not pink and not in my field of vision might have been pink and in my field of vision? After all, the rat might have been green and next door. Why, then, do I experience a pink quale, rather than a green one or none at all? The possibility that a particular absent grey thing might have been present and pink seems too tenuous a state of affairs to bear the explanatory burden that the relocationist$_w$ wants to place on it. The mechanism that might underwrite the causal connection between me and my intentional object so specified remains hidden. But without specification of this machinery the claim to have solved the causation-problem remains hollow.

This reading of the ersatzist position bears some similarity to Lewisian realism. The intentional objects that bear the phenomenal pink are neither Brentano's mental objects nor Meinong's homeless objects. They are physical objects much like the ones we interact with every day. The robustness of the objects involved allows these theories to deal successfully with the nature-problem. But the two positions pull apart when we look at the reasons for their inability to solve the causation-problem. Lewis's pink rat cannot account for the hallucinator's experiencing a pink quale because Lewis's pink rat is in a world that is sealed off causally from our world. The ersatzist's pink rat cannot account for the hallucinator's experiencing a pink quale, because in the actual world, where the rat's color could cause one to experience a quale, the rat is grey. To say that there is a world in which this very rat is pink and could, therefore, cause one to have a pink quale does not help. For that merely amounts to saying that this rat could have been pink. It must, of course, be granted that the grey rat could have been pink. But this is not the kind of fact that one can appeal to when one endeavors to explain why a person is actually experiencing a pink quale in a situation in which the relevant possibility is not fulfilled.

Our brief tour through the various ontologies of merely intentional objects has discovered nothing that might help the case of relocationism$_w$. First we found that neither Brentano's mental objects nor Meinong's homeless objects are suitable to exemplify the qualia we have in nonveridical experience. Then we saw that the attempt to understand merely intentional objects in terms of possible worlds is unsuccessful. Neither Lewsian realism nor ersatzism have proven workable. This brings up to four the number of failed attempts to provide the WRT with a suitable intentional objects to bear the qualia of nonveridical experience. Perhaps there are other avenues for the relocationist$_w$ to explore. But by

now it should be clear that the difficulties posed by nonveridical experience are formidable.

9.2.4 How to Perceive Otherworldly Objects

In Chapter Eight I made much of the difficulties that are involved in the having of the relocated qualia of veridical experience. Now I have argued that the case of nonveridical experience compounds these problems. Exiling qualia of nonveridical experience into neighboring worlds guarantees that they do not cause ontological trouble in our world. It also seems to guarantee that they do not cause anything in our world. The first consequence is what makes the deportation of nonveridical qualia attractive. The second (is part of what) makes it objectionable. The causal isolation of mere possibilia makes the causation-problem—and therefore also the having-problem—insoluble.

This objection is not new to those who enlist mere possibilia as intentional objects of nonveridical experience. Lycan, for example, has seen it and brushed it aside:

> If there were really sense-data, with which we were acquainted, there would be a serious question as to *how* we could be acquainted with them given that the notion of acquaintance is a causal one. The present idea avoids that question. To be "acquainted with" a nonveridical "sense-datum" is to represent a nonexistent object. (Lycan 1987: 89)

Note, first, that Lycan does acknowledge a causal constraint on a naturalistically respectable account of the having-relation. In the first part of this passage he pursues precisely the course that I have been following in this discussion: he argues that the causation-problem is a serious difficulty for the dualist. I agree. But then he turns around and asserts that the causation-problem simply does not arise for those qualia-bearers that are mere possibilia. If this is so the argument that the causation-problem for mere possibilia is insoluble is pointless. The causation-problem need not be solved in this case because it does not even arise.

What is one to make of this reply? On the relocationist$_w$ picture, qualia are external and *they*—these external qualia and not some inner representatives of them—suffuse our experience. When you look at a ripe tomato its redness is the redness that characterizes your experience. There is only one redness in play and it does double duty, so to speak. It makes the tomato red and your experience a red-experience. We have seen various representation-based accounts of how this is supposed to happen. Leaving aside the question how successful they are in explaining this transaction, we can observe that these accounts have one thing in common. They identify the external color property of your intentional object with the color quale you enjoy in the experience, and they hold that representa-

tion is the mechanism that makes this possible. In representing the external property, we make it our own. It is as if the act of representing the tomato opened up a window onto the tomato, thereby letting its redness color your experience. For that to work, the red has to be there. Opening up a window onto nothing won't impart much color to your experience. For there is nothing there to color it with. Lycan's sense-data are nothing—they do not exist. How is representing them going to impart any phenomenological dimension to your experience?

For this reason I am not satisfied with Lycan's way out of the causation-problem. When considered from within a naturalistic framework that combines relocated$_w$ qualia with a representation-based account of how those qualia are had, the act of representing a nonexistent object should be like nothing for the representing subject. Thus, to put it in Lycan's terms, there still is a serious question as to how we can be acquainted with the nonexistent possibles that bear our nonveridical qualia.

That is not to say that nobody can help himself to Lycan's answer. It all depends on your other philosophical commitments. If, for example, you hold that qualia are mental paint, then your story about what it is to represent a nonexistent object will be very different. Roughly, it will be a matter of your tokening an inner representation with a full set of qualia and then using it in the appropriate manner or making it the object of a HOR. If this representation does not correspond to any external object, what you will have done is to represent a nonexistent object. Doing so will, nevertheless, have been like something for you, because the qualitative feel of the experience is due to the activity in your head which can unfold flawlessly even if it corresponds to nothing at all in the world that surrounds you.

Or consider a theory of the mind that is not bound by naturalistic constraints. For concreteness's sake, think of Moore's account of sensation according to which the mind reaches out and knows things, where knowing is a "simple and unique relation" that characterizes the mind and differentiates it from matter. On such a theory, the mind may be capable of feats that no ordinary mechanism could duplicate. A mind, so conceived, might reach into the possible world where Lycan's nonveridical sense-data reside, making their acquaintance in the process.

But Lycan will have neither mental paint nor paramechanical minds. Abstaining from these props, and trying to get by with physical paint and naturalistically respectable mechanisms, is commendable but hard. For it rules out the simple answer to the question how we get to have otherworldly qualia. Given this sparse framework, it will not do to say that having the qualia of nonveridical experience is simply a matter of representing the nonexisting sense-data that bear them. Doing that would be like nothing. That is, doing that is not a way of hav-

ing the qualia that nonveridical experience affords us. Thus, the simple answer to the question how we get to have otherworldly qualia is ruled out. The causation-problem posed by nonveridical experience still stands.

9.2.5 Halfhearted Relocationism and the Problems of Illusion

The WRT was the target of the preceding discussion of the problems arising from nonveridical experience. The HRT made only one brief appearance and then only to be favorably contrasted with the WRT (see page 214). I proceeded in this way because it is obvious how these problems arise for the WRT. I will now briefly argue that the same sorts of problems arise for the HRT in a less obvious but no less damaging way.

Above I provided a brief and schematic account of the way in which the HRT can accommodate nonveridical experience. This may have created the impression that the mental-paint story is much superior to the physical-paint story when it comes to accounting for our ability to represent the nonexistent. But the situation is less clear-cut than it appears on first sight. Problems similar to those that plague the physical-paint account can arise within the mental-paint setting. This is the case if a higher-order representation is directed at a lower-order mental state that does not exist. Rosenthal, for example, acknowledges that this is a possibility and he adds that "a case in which one has a HOT along with the mental state it is about might well be subjectively indistinguishable from a case in which the HOT occurs but not the mental state." (Rosenthal fc.: Section V) The problems this raises are precisely analogous to the ones we encountered in discussing Lycan's attempt to come to terms with nonveridical experience. In these cases, the qualia-bearing mental states do not exist, yet we are (transitively) conscious of them and they thereby become (intransitively) conscious. Thus, the merely intentional objects are back. This time around they are mental states rather than external physical objects. But this is a difference that makes no difference. The exemplification-problem and the causation-problem arise in the same way for both these kinds of nonexistent possibles.

The specific problems that were obvious in the case of the relocation$_w$ scenario are no longer visible in the picture that the HRT paints. But embracing the mental-paint account does not make these sorts of problems go away. It merely pushes them back one step by moving them inside where they are concealed from clear view.

9.3 The Argument from Atomism and Illusory Qualia

I have argued that nonveridical experience poses a formidable problem for the relocation theses. But perhaps nonveridical experience does not deserve that much attention. After all, these instances of nonveridical experience constitute only a minority of corrupt cases within the vast number of veridical experiences. In fairness to the relocation theorist, we should test his theory in the field of veridical experience, where we have real physical things to support the phenomenal properties. If he succeeds here, he will have succeeded in making an impressive case for his thesis. Conversely, if he fails here, his thesis will have nothing left to recommend it. The argument from atomism attempts to show that the relocation theses are untenable even in the case of veridical experience and that they must, therefore, be given up.

Simply put, the problem is this: qualia (as traditionally conceived) don't agree with matter. The more we learn about matter, the less hospitable does it appear to phenomenal properties.[123] From its inception up to the present time, atomism (or the corpuscularian philosophy) has presented a picture of matter that purportedly excludes qualitative properties. It all started with Democritus who demoted sensory qualities to merely conventional items that form no part of reality (see page 2), and the view is still well and alive in our century where it found its most famous expression in Arthur Eddington's claim about the two tables. There is the familiar table of everyday life, which Eddington characterizes as extended, comparatively permanent, colored, and substantial. (See Eddington (1929: ix)) And then there is the scientific table: "My scientific table is mostly emptiness. Sparsely scattered in that emptiness are numerous electric charges rushing about with great speed; but their combined bulk amounts to less than a billionth of the bulk of the table itself." (Eddington 1929: x) Although Eddington does not explicitly assert that the two pictures are incompatible, this seems to be implied by what he goes on to say. He tells us that "modern physics has by delicate test and remorseless logic assured me that my second scientific table is the only one which is really there—whatever 'there' may be." (Eddington 1929: xii) Thus the ordinary table of everyday life—"strange compound of external nature, mental imagery and inherited prejudice" (Eddington 1929: xii)—loses its status as a real thing. And towards the end of his investigation Eddington informs us that "our conception of the familiar table was an illusion." (Eddington 1929: 323)

Clearly, Eddington's assertions could be reconstrued as an argument to the effect that a realistic interpretation of physical theory forces us to give up our commonsensical view of ordinary physical objects. Tables—i.e., real scientific tables, which are the only tables that genuinely exist—are not colored. Science teaches us that real tables cannot support phenomenal qualities. Thus, the attempt

to relocate$_w$ the color qualia I experience upon viewing a table on the table itself must fail. Due to Wilfrid Sellars's tireless efforts to clarify the relationship between the manifest and the scientific image of man-in-the-world, different and much refined versions of the argument from atomism—or the 'grain argument,' as he called it—are now in existence.[124] Here is how Sellars formulates the intuition that the grainy structure of matter is in conflict with the homogeneity of, for example, colors as we experience them:

> The 'ultimate homogeneity' of perceptible qualities which, among other things, prevented *identifying* the perceptible qualities of physical objects with complex properties of systems of physical particles, stands equally in the way of *identifying*, rather than *correlating*, conscious sensations with the complex neural processes with which they are obviously connected. (Sellars 1963: 36)

So much for a sketch of the argument from atomism and of how it is supposed to undermine the relocationist move. Here I will not discuss the question whether the natural philosophers from Democritus to Eddington and Sellars have been right to insist on the genuineness of the grain problem that underlies the argument form atomism. I avoid the question whether, and if so how, the manifest and the scientific images clash—not because of a lack of intrinsic interest of this issue, but simply because such a discussion cannot be carried out profitably without a solid understanding of physical theory, which I lack.[125]

Instead, I will focus on one aspect of a philosophical theory that attempts to defend the relocationist thesis in what I take to be a most original way. The defense relies on a strategy that atomists of all ages apparently failed to consider. One thing that stood fast for these thinkers was the claim that phenomenal qualities are simple, unanalyzable, and completely known to the person who experiences them. The thesis under discussion denies this more or less explicitly held assumption that has governed all thought on this issue. In making the switch from silver-screen to celluloid qualia contemporary qualophiles have turned qualia into ordinary, naturalistically respectable properties. Celluloid qualia have none of the 'ultimate homogeneity,' simplicity, irreducibility, etc., of their traditional forebears. The graininess of matter does not prevent the identification of celluloid qualia with physical or neurophysiological properties. Thus worries about grain and the argument from atomism are a thing of the past. But there is a dark cloud on the horizon. For even those who take qualia to be complex physical properties of some sort must admit that that is not the way our qualia seem to us. When confronted with this challenge they reply that the apparent simplicity and homogeneity of our qualia is illusory. This appeal to illusion is what I want to focus on in the following discussion. I will argue that everything about this purported illusion is problematic. The nature of this illusion remains obscure.

The attempts to account for the origin of the illusion do not satisfy. And the extent or the depth of the illusion must be much greater than we are given to believe. To speak of an 'element of illusion' in this context is altogether too moderate. If this account of phenomenal qualities is correct our experience of them is illusory through and through. The illusion you suffer when you see a red spot is no less complete than the illusion experienced in after-imaging a red spot. In both cases, the red spot, as you experience it, is not there in front of you. Thus, the problems that beset nonveridical experience also pose themselves in the domain of veridical experience.

9.3.1 Direct Reductive Realism about Color

I begin by placing this new line of argument into a broader context. Focusing on phenomenal color, the position we are considering can be characterized as a form of color realism. Realism, in this context, is to be understood as the claim that the existence and nature of colors is independent of their being experienced. The red I currently experience—say, the redness of the ripe tomato I am looking at—would be what it is, even if I had not seen it now, even if I had never seen it, even if nobody had ever seen it, even if nobody ever could have seen it. The particular version of realism that interests us is a form of direct realism for it claims that we gain awareness of colors noninferentially and unaided by any intermediary entities, like, e.g., the sense-data of the representative theories of perception. As characterized so far, the position we are looking at does not involve anything extraordinary: naive realism—allegedly the belief of philosophically unenlightened masses—is just such a form of direct realism. What sets this position apart from a naive form of direct realism is the claim that the directly apprehensible, mind-independent phenomenal color properties are reducible to complexes of primary properties. What we are dealing with, then, is a form of direct reductive realism about color, or 'DIRE realism' for short. Armstrong, the most tireless defender of DIRE realism, has succinctly characterized the reductive component of this doctrine: "Secondary qualities are primary qualities not apprehended as such." (Armstrong 1984: 185)

The DIRE realist acknowledges that the argument from atomism is a powerful objection to non-reductive forms of direct realism. In fact, Armstrong seems to grant that the reducibility of secondary qualities to primary properties is the touchstone of physicalism:

> Suppose, however, that the secondary qualities are irreducible. The same will have to be said, presumably, about the qualities associated with bodily sensations and perhaps other mental processes, such as the emotions. The body and the mind will then have associated with them a range of qualities not to

> be found in ordinary physical objects. There will then be a case for thinking that body and mind are special sorts of objects, set apart from ordinary material things. For a materialist, then, much turns upon whether he can give a satisfactory materialist theory of the secondary qualities. (Armstrong 1984: 185)

If the atomists are right about the kinds of properties that can be ascribed to matter, then finding a home for the irreducible phenomenal properties will be problematic. Neither the physical objects to which we are related in experience nor our brains (or bodies) are the sort of thing that can host such irreducible qualia. If we continue to hold fast that such irreducible phenomenal properties exist, then the atomistically conceived world will have to give somewhere: something in it must be such that it can bear these unruly properties. This is the route leading from the argument from atomism, plus the assumption of the irreducibility of qualia, to the abandonment of physicalism.[126] Giving up the irreducibility of phenomenal properties is an ingenious move on the part of the DIRE realist. It enables him to achieve two goals at once. First, it makes it possible for him to acknowledge the power of the argument from atomism in the intellectual context in which it was proposed. This is a welcome consequence. For, if you claim the untenability of an historically important argument, it is incumbent upon you to explain the argument's seductive power over the minds of your predecessors. By granting the argument's soundness, given the historically prominent irreducibility assumption, the DIRE realist does explain the argument's robust appeal. Second, in giving up the irreducibility assumption, the DIRE realist eliminates the principled opposition between primary and secondary qualities. Collapsing this opposition renders the atomist's insistence that matter has (at best) only primary properties harmless. If secondary qualities are primary properties, then matter can have secondary qualities as well.

9.3.2 The Illusion of Concrete Secondary Quality: Armstrong's Account

Now that we have a better understanding of DIRE realism and the importance it attaches to the reducibility of phenomenal qualities, I can finally state the problematic corollary of DIRE realism. Armstrong admits that there is a price to pay for the reduction of secondary qualities: the DIRE realist has to admit that *"sensory illusion is a much more widespread affair than we ordinarily assume."* (Armstrong 1961: 164) Writing sixteen years later he acknowledges that, on his theory, the "perception of secondary qualities involves an element of illusion," (Armstrong 1977: 31) and still more recently Armstrong wrote that his view has the consequence that we suffer from "the illusion of concrete secondary quality." (Armstrong 1984: 180) There can be no doubt that the appeal to illusion is a crucial and enduring trait of Armstrong's DIRE realism. DIRE realism would be in

bad shape if it were to turn out that the appeal to the illusion of secondary quality is riddled with problems.

Armstrong's first pass at illuminating the nature of the illusion reads as follows:

> When in perception we are aware of the color, sound, taste, etc., of the physical things, then the qualities which we are aware of are complexes of physical properties. The perceived secondary qualities are primary qualities! But we are aware of them in a unified, *Gestalt,* manner, a manner which fails to reveal the primary nature of these properties. (Armstrong 1984a: 178)

Our experiential grasp of the secondary qualities is misleading: we experience simplicity and homogeneity where there is structural complexity. Take, for example, the secondary quality of phenomenal greenness: the quality as we experience it and the property as it really is differ significantly. In what sense, then, are we dealing with the *one* property of phenomenal greenness? This difficulty has not gone unnoticed:

> When a surface looks green to me, I am not attributing the corresponding primary quality or, indeed, any primary quality. Yet I am attributing some property to it. So am I not attributing a property to the surface which is *different* from the corresponding primary quality? (Armstrong 1984: 179)

This makes DIRE realism look bad. The talk of illusion harbors unforeseen dangers. Instead of the intended reduction of ontological baggage, DIRE realism seems to have led to an increase in the number of properties. In addition to primary properties and secondary qualities, the DIRE realist needs to countenance properties that might be called "primary property counterparts of secondary qualities"—properties that are constituted by clusters of primary properties. Whether the acknowledgment of these counterpart properties will actually lead to an inflation of the realm of properties will depend on one's manner of counting properties. But whichever way one counts, it is certain that DIRE realism has not achieved its goal of increasing ontological economy.

In order to make some progress, the DIRE realist must find a way to identify secondary qualities with their primary property counterparts. Armstrong attempts to facilitate this identification by radically overhauling our notion of secondary qualities. Employing the "topic-neutral maneuver," he proceeds to deplete phenomenal properties of all content. He suggests that we should think of secondary qualities as "*qualities we know not what.*" (Armstrong 1984: 180) Thus purged of what we took to be their intrinsic nature, secondary qualities will not resist identification with whatever our science deems the most plausible candidate. This treatment of secondary qualities closely parallels the strategy Armstrong used on mental states. The concept of a mental state is the concept of a state apt

to bring about certain effects. A mental state concept of this kind does not tell us anything about the intrinsic nature of the state. This is the sense in which the mental state concepts contain a blank or gap. It takes scientific investigation to find out how this gap is to be filled. Armstrong alludes to this parallel when he writes:

> Suppose that our concept of red is *all* blank or gap? May it not be that we know *nothing* about what redness is in its own nature? May it not be that we only know contingent truths about redness—such truths as that it is a property detected by the eye and possessed, or apparently possessed, by such things as the surface of ripe tomatoes and Jonathan apples? Then it would be possible to go on to a contingent identification of redness with a physical property of the red thing. (Armstrong 1968: 275)

As we have already seen in Section 3.1, Armstrong is well aware that this account of secondary qualities faces "an enormous phenomenological difficulty." (Armstrong 1984: 179) For in experience we seem to become fully acquainted with the striking intrinsic natures of these phenomenological properties. This implausibility of DIRE realism can be mitigated if it is possible to show that this seemingly unqualified grasp of the intrinsic natures of secondary qualities is really an illusion.

Armstrong relies on three considerations to unveil the illusion of concrete secondary quality. We already encountered his idea that we grasp secondary qualities in a *Gestalt* manner. He supplements this consideration with two others. First, we suffer from an illusion that is analogous to the headless-woman illusion. Take a stage with a black background on which stands a woman whose head is draped in black. By rendering the woman's head invisible, this arrangement creates the viewer's illusory impression of seeing a headless woman. In parallel fashion, our failure to detect any complexity in the phenomenally given secondary qualities creates the illusory impression that we notice the simplicity of these properties. The second and more important consideration is based on our awareness of the complex and systematic interrelationships obtaining between the secondary qualities, especially the colors. Unaware of the real primary-property basis of these resemblances and differences between colors, we erroneously take ourselves to have an awareness of the intrinsic natures of secondary qualities, because

> given our instinctive taking of resemblance to be an internal relation, a mere perception of resemblance suffices to generate the illusion that we have a concrete acquaintance with the qualities which sustain the resemblance. *A perception of the internal relation of resemblance generates the illusion of a perception of intrinsic quality.* (Armstrong 1984: 180-181)

Now all the pieces of the picture are in place. The correct account of our notion of secondary qualities reveals them as mere 'that whiches' whose internal nature is unspecified. Scientific considerations lead us to identify them with complexes of primary properties. Our mistaken impression that we are very well acquainted with the intrinsic nature of the secondary qualities derives, for the most part, from our inability to tolerate brute similarity relationships between secondary qualities. The true grounds of these relationships are phenomenally inaccessible to us. For want of a correct explanation, we embrace the reassuring illusion that we have an intimate acquaintance with the qualitatively simple natures of the secondary qualities.

9.3.3 Piercing the Illusion

Armstrong's defense of DIRE realism is impressive. Nevertheless, I find one of its central features puzzling and obscure. This feature is the DIRE realist's appeal to illusion. Is the appeal to illusion an indispensable feature of DIRE realism? I believe it is. The appeal to illusion is what gives DIRE realism whatever measure of phenomenological plausibility it has. Stripped of its illusory element, DIRE realism would no longer have a place for the qualia of our experience. It would, in short, collapse into something akin to qualia eliminativism.

When confronted with the patent difficulty involved in identifying color qualia with primary-property complexes, Armstrong makes two coordinated moves: the topic-neutral maneuver and the appeal to illusion. Both moves are indispensable components of DIRE realism. By employing the topic-neutral maneuver Armstrong strips the phenomenal properties of their identification resistant features. Mere 'that whiches' have no backbone; the theoretician need respect no constraints when he sets out to discover the true nature of such blanks. Thus, the task of identification becomes easy. But by turning phenomenal properties into complaisant properties, the topic-neutral maneuver has robbed those properties of their most interesting feature: their phenomenal character. What would it be like to have topic-neutral 'that whiches'? If red confronted us as 'all gap,' why should it be like anything to have this quale? It goes to Armstrong's credit that he notices and takes seriously this disastrous consequence of the topic-neutral maneuver. It is the enormous phenomenological difficulty of which he speaks. Had he ended his account of DIRE realism at this point, the theory would be guilty of what we might call 'phenomenology neglect'—a complete unawareness, and therefore a complete failure, to address the phenomenologically salient features of experience. While not technically a form of qualia eliminativism, a neglect theory of this sort would hold little interest for one interested in qualitative consciousness.

This is where the second move—the appeal to illusion—comes into play. Its purpose is to restore the phenomenological adequacy of the DIRE-realist account. The phenomenal aspect of phenomenal properties (which was banished by the topic-neutral maneuver) reenters the picture in the guise of an illusory component or our experience of phenomenal properties. Everybody is supposed to be happy. The qualophile has been provided with his warm phenomenal glow upon which she insists, and the scientifically minded reductionist has been furnished with well behaved phenomenal properties that do not buck identification with scientifically respectably properties. This, then, is the role of the appeal to illusion in DIRE realism: it prevents the theory from being glaringly inadequate on the phenomenological front. And now it should also be evident why the appeal to illusion is of crucial importance. Inasmuch as one is interested in an explanation rather than an elimination of qualia, the appeal to illusion is an indispensable part of the DIRE-realist project.

In what follows I argue that Armstrong's appeal to illusion is problematic. First, I argue that it is surprisingly hard to get a clear notion of the nature of the illusion involved in the illusion of concrete secondary quality. Second, I question Armstrong's account of the origin of the illusion. Lastly, I argue that the extent of the illusion is far greater than the DIRE realists would have us think. Our experience of phenomenal color does not merely contain an element of illusion; it is wholly illusory. It follows that the DIRE realist's account of veridical color experience inherits all the problems that beset the account of nonveridical experience. Hence, the attempt to save DIRE realism by demonstrating its power in its chosen field of veridical experience must be regarded as a failure.

9.3.4 The Nature of the Illusion

How are we deceived when a tomato looks red to us in the customary way? What is the error involved in this perception? On the old sense-datum theory these questions were not hard to answer. According to this view, the colors exemplified by sense-data differ profoundly in nature and location from the colors exemplified by the external physical objects. Even after one learns this fact one tends to slip back into the naive view attributing the inner colors to the outer objects. Thus one mistakes the nature of the outer colors by identifying them with the inner colors and one mistakes the location of the inner colors by projecting them outward. Given this account of the nature of illusion, Armstrong's attempt to enlist the machinery of illusions to account for the troublesome homogeneity and simplicity of experience must appear quite wrongheaded. For, on the sense-datum account, the illusion of concrete secondary quality consists precisely in my being acquainted with the simple, homogeneous redness of a sense-datum.

The ease with which the sense-datum model explains our mistaken belief that the colors of physical objects are simple and homogeneous stems from the fact that it recognizes to sets of colors—the colors of physical objects and the colors as they are exemplified by sense-data. If one works with two sets of colors it is easy to make mistakes and it is easy to explain how these mistakes are made.

This consideration shows why Armstrong's appeal to illusion may seem problematical. The traditional account of the illusion of concrete secondary quality helps itself to two sets of colors. Armstrong acknowledges the existence of only one set of colors. Moreover the second set of colors that the sense-datum theory works with are precisely the sorts of properties—simple, homogeneous qualia—that Armstrong wants to avoid by appealing to the illusion of concrete secondary quality. Thus Armstrong owes us an account of the illusion that does not rely on the problematical properties that the appeal to illusion is supposed to dispense with.

Armstrong cannot use the sense-datum model of the illusion because he works with only a single set of colors—the physical colors of external physical objects. Yet, he insists that we make mistakes and acknowledges that he has to explain how they are made. So, how is he going to account for the tomato's looking simply and homogeneously red? Unlike the sense-datum theorist, he has not available a simple homogeneous red quale the having of which might constitute this seeming. The only red he has available—the only red that he could offer as the quale of our tomato experience—is the primary-property complex residing on the tomato's surface. But all parties to the dispute agree that this is not the way tomatoes look to us. That tomatoes do not look as if covered by a primary-property complex is precisely what gave rise to the problem of the illusion of concrete secondary quality in the first place. In trying to come to terms with the nature of the error involved in the illusion of concrete secondary quality, we have come back to the problems raised by the argument from illusion (see especially Section 9.2.4). It seems that the relocationist who is also a reductionist about color cannot provide a quale in terms of which the tomato's looking homogeneously red can be explained. But note that this time the problem arose within the domain of veridical experience. The tomato that looks red is really there and really red. Yet this real redness, the only property in the situation positioned to serve as the quale of the observer's tomato experience, cannot do the job. How, then, does the tomato manage to look red in the familiar homogeneous and simple way?

9.3.4.1 Seeming and Believing

Armstrong's account of perception suggests that the illusion of concrete secondary quality is grounded in beliefs about simplicity of qualia and not in simple

qualia. This allows him to acknowledge that the experience of color seems to involve direct acquaintance with simple color, while holding that nothing exemplifies simple colors.

A subject who reports that the tomato seems simply and homogeneously red to her, it might be argued, is not issuing a pure and unadulterated phenomenological report. Rather, the subject who tells us how it seems to her is telling us something about a belief of hers. That is, the simplicity and homogeneity of a red experience is not part of the phenomenological core of the experience; it is a belief that is superadded to the experience. Perhaps it is something you cannot help but believe, given the kind of being you are and the kind of stimulation you receive. On this account, representing a complex property as simple is not a matter of your being acquainted with a simple property. Nor is it simply a matter of your being acquainted with a complex property. Representing a complex property as simple is a matter of your being acquainted with a complex property and of your tokening certain erroneous beliefs to the effect that the object of your acquaintance is simple. Thus the illusion of concrete secondary quality is traced back to beliefs about the simplicity of certain things, rather than to the simplicity of those things. The detailed explanation how representing real complexity results in apparent simplicity and homogeneity will therefore have to account for the origin of these mistaken beliefs about simplicity. Armstrong's account of the origins of the illusion is the subject of Section 9.3.5.

How should one assess the doxastic account of the illusion? Keep in mind that the illusion of concrete secondary quality is of a peculiarly phenomenological kind. That is, the belief that the experienced red is homogeneous and simple seems to be underwritten by a phenomenological datum. Unlike the belief in the simplicity of God, the soul, and the monad, the belief in the simplicity of the red quale appears to be a deliverance of experience. Any adequate account of the illusion of concrete secondary quality will have to do justice to this phenomenological dimension. It will, therefore, not do to say that *all* that goes wrong is our erroneous belief in the homogeneity and simplicity of the tomato's redness. After all, a blind person can believe that, yet the blind person is incapable of being phenomenologically misled in the way we are. To her the tomato does not look to be coated with the homogeneous and simple redness that we enjoy. But the tomato's looking that way to us is, of course, at the very heart of the illusion we suffer. Thus, the attempt to provide a belief-based account of the illusion of concrete secondary quality strikes me as inadequate. If the relevant beliefs can be tokened without the mobilization of a simple and homogeneous quale, they will fail to capture the phenomenologically striking feature of the experience of redness. And if they capture this feature, then they will involve the sort of problematic item they were supposed to dispense with.

Why, then, is it so difficult to rid oneself of the feeling that there must be something right about this doxastic approach to the illusion of concrete secondary quality? I believe that this impression is grounded in a certain shiftiness of the terms like 'seeming' and 'appearing' and their more specific cognates—'looking,' 'sounding,' 'smelling,' etc. These terms have a number of quite distinct uses. The most relevant for us are the doxastic use and the nondoxastic, descriptive or phenomenological, use. In their doxastic use the appear-words express what the speaker believes. In their descriptive or phenomenological use they do not express beliefs; they report the intrinsic nature of the speaker's experience. Much more can be (and has been) said about how they manage to do the latter.[127] But even without the benefit of this discussion it is easy to find examples in which appear-words function nondoxastically. Take, for example, Chisholm's sentence: "That thing appears white to me in this light, but I know that it is really grey." This sentence makes sense only if the term 'appear' is not taken doxastically. Not paying sufficient attention to this distinction between doxastic and nondoxastic uses of appear-words makes many problems appear simpler than they are. This manner of 'simplifying' problems is most easily accomplished using the term 'seems.' Called upon to defend the phenomenological adequacy of one's account, the word 'seems' obligingly slides towards the phenomenological end of the spectrum. But by the time one gets around to providing an account of the way things appear to us, the word 'seems' has bounced back all the way to the belief-end of the 'seeming-continuum,' thus smoothing the way for a doxastic account of the phenomenon. So the impression is created that all is well with the illusion of concrete secondary quality—without compromising the constraint of phenomenological adequacy, we have arrived at a lucid and nonmysterious account of the illusion. Alas, the impression is an illusion fostered by the unrecognized shiftiness of the notion of seeming.

9.3.4.2 Seeming and the Language of Thought: Lycan's Way Out

I have chosen Armstrong to be my guide through the many problems that atomism raises for the WRT. But now I want to briefly abandon Armstrong to see whether Lycan can offer a more satisfactory account of the illusion of concrete secondary quality.

Lycan has refined HOP by combining it with a language of thought account of representation. He argues that this augmented version of HOP provides plausible answers to a number of problems usually (but inaccurately) lumped together under the label 'the problem of consciousness.' Nagel's "intrinsic perspectival, point-of-view and/or first-person aspect of experience," the "funny facts" discovered by Frank Jackson's Mary, and Joseph Levine's "explanatory gap" are three of the problems he mentions. (See Lycan (1996: 9)) Here I look at

Lycan's attempt to come to terms with the explanatory gap. For it seems that his treatment of this problem points to a promising, representation-based, answer to the having-problem. More particularly, Lycan's approach seems to offer a phenomenologically adequate answer to the question why relocated celluloid qualia—the complex, structured primary quality complexes of the WRT—appear simple and homogeneous to us.

Let us begin with Lycan's presentation of the problem of the explanatory gap:

> The token-identity thesis does nothing to *trace* the relation of sensations to neural firings, and it does nothing to explain why sensations have the particular feels they do, even given its philosophical tenability. In these senses, materialism has failed to *account for* consciousness...(Lycan 1996: 63)

The augmented HOP suggests the following solution to this problem:

> the lack of tracings and explanations of the sort demanded is *just what you would expect* if the [augmented] self-scanner view of introspection is correct...therefore, the lack of such tracings and explanations, only to be expected, do not count against the materialist identification. They almost count in its favor. (Lycan 1996: 64)

Lycan's strategy for saving materialism does call to mind a favorite ploy of the software industry: a program's bugs are 'fixed' by calling them features. But here I do not want to pursue the question whether this answer succeeds in assuaging the worries raised by the explanatory gap. Instead, I want to look at the explanatory apparatus Lycan uses to 'bridge' the explanatory gap and ask whether it provides an account of the illusion of concrete secondary quality.

To introspectively represent a mental state is to mobilize a term of the language of thought (also called 'Mentalese') that stands for that state. Lycan notes two special features of these Mentalese tokens. They are simple expressions of the language of thought and they play a unique functional role in the mental economy of their user. Thus we get the result that such an introspective representation "would be a private name as well as semantically primitive, a name that only its actual user could use to name its actual referent." (Lycan 1996: 60) The combined features of primitiveness and privacy of the Mentalese words I use to introspectively represent my mental states make it impossible for anybody else to represent my mental states in a relevantly similar way. Others can think of my mental state and they can speak about it. But they cannot do it under the same mode of presentation. Here is how Lycan puts it:

> And since no one else can use that mental word or even any functionally or syntactically similar word of their own to designate that state of affairs, of course no one can explain in English or in any other language why that state of affairs feels like *that* or *semantha* to me. (Lycan 1996: 64)

Let us grant, for the sake of the argument, that this does explain why the explanatory gap opens up. But does the appeal to the primitiveness and privacy of the representation I token to introspect my tomato perception explain how I get to *have* the red quale that the introspected perceptual state represents? Does it explain why there is something it is like to be the one who represents the tomato perception in this way?

It may seem that the appeal to primitiveness and privacy can speak to these questions. This becomes more readily apparent when we turn to a slightly more specific question: Does the appeal to the primitiveness and privacy of the representation I token to introspect my tomato perception explain such salient features as the apparent simplicity of the red quale I have in seeing the tomato? The answer to this question may seem to be close at hand. The represented quale is complex. But the representation by means of which the perception of the quale is introspected is simple. Therefore, the resulting experience is as of a simple quale. The simplicity belongs to the representing, not to the represented. So the impression of the quale's simplicity is explained and the source of the illusion of concrete secondary quality stands revealed.

The ring of plausibility this account enjoys is due to the impression that it takes features the experienced quale seems to exemplify and attributes them to the machinery that allows one to have the quale. This is the familiar pattern of explanation found in the sense-datum theory. Sense-data are said to really have the properties that the perceived object seems to have. Undoubtedly there are also significant differences between Lycan's Mentalese terms and sense-data as traditionally conceived. Notwithstanding these differences, the two theories are seen to have significant parallels when viewed from a sufficient distance. Both theories work by reassigning problematical properties to a new bearer. When an object appears to have a property that it cannot have, both these theories instruct us to reassign the offending property to the token we use to represent the object. Thus problematic properties ostensibly characterizing the object turn into well-behaved properties of the object's representation. Failing to notice this displacement of properties, we mistakenly continue to experience those properties as belonging to the experienced object.

When one sees a ripe tomato, for example, the sense-datum theory assigns the simple redness one experiences to the sense-datum. On this account, there is no mystery why tokening *such* a sense-datum would lead one to experience the tomato like *that*. The properties indicated by the 'such' and the 'that' are identical—a simple red quale. Things are not similarly straightforward in the case of the augmented HOP. The simple redness the tomato seems to exemplify does not migrate into the head to alight on a token of Mentalese. Nothing in the head is red in the way in which the tomato seems to be red. Moreover, the presence of

such an item in the brain would not explain how the brain's owner gets to have the quale exemplified by his brainstate. The gross image of Nagel's gourmet scientist vividly makes this point (see Section 7.5.3). The DIRE realist about color holds that the redness we experience in looking at the tomato is outside and stays there. Not the tomato's redness itself but merely some of the salient features that characterize this redness are moved onto an item in the head. Only the apparent simplicity of the tomato's redness—not the redness itself—is a real feature of the Mentalese term we use in introspection. Thus, the feature shared between the way the tomato looks and the representation one tokens in introspecting one's tomato perception is quite abstract.

This constitutes a significant deviation from the model provided by the sense-datum theory. The sense-datum theory takes the property that characterizes what it is like for you to see a tomato and attributes it—all of it—to your sense-datum. The augmented HOP theory takes this same property, abstracts a certain higher-order feature from it, attributes only this higher-order property to your introspective representation, and leaves the remainder of the original property on the tomato's surface. Phenomenal redness has thus been bifurcated into two components, each of which is phenomenologically elusive. We never experience the real, external redness as it is in itself. And the illusory simplicity contributed by our representational machinery is hard to picture abstractly—detached, that is, from simple, homogeneous red, or green, or blue, etc. How these two distinct components—the real external redness and the illusory simplicity—coalesce to form the impression of the seamless quale that characterizes our experience, is difficult to comprehend.[128]

But even when we restrict our attention to the dimension of simplicity alone, the augmented HOP encounters difficulties not found in the sense-datum theory. For it appears that there is no *one* property that a simple quale and a simple lexeme of Mentalese have in common. The simplicity of the former is phenomenological; that of the latter semantic. The analysis of phenomenological simplicity, and the homogeneity that comes with it, is "extraordinarily complicated." (Lycan 1995: 8) The analysis of the semantic simplicity involved in lexical semantics poses no parallel problems. To this we might add the observation that phenomenological simplicity seems ultimate in a way that lexical simplicity is not. The analogy between natural language and the language of thought suggests that the lexical level of individuation of Mentalese items is not fundamental. Morphology and phonology provide manners of semantically individuating linguistic items on a more fine grained scale than lexical semantics. What is simple on the lexical level may appear complex from the phonological or morphological level. But we find no parallel relativity of phenomenological simplicity. You cannot, for exam-

ple, use a microscope to discover new layers of phenomenological simplicity more fundamental than those encountered in unaided vision.

These considerations suggest that the linguistic representations allegedly mobilized in introspecting one's perceptual states have none of the salient qualities that the objects of our experience seem to have. The Mentalese token used to introspect one's perception of the ripe tomato is not red in the way the tomato seems red, nor is it simple in the way the tomato's redness seems simple. Thus, the augmented HOP theory inherits none of the plausibility of the sense-datum theory's answer to the question why tokening *such* a sense-datum would lead one to experience the tomato like *that*. The sense-datum theory answers that experiencing the tomato is like *that* because *that* is the quality your tomato sense-datum really has. The augmented HOP cannot give an analogous answer, because *that* is not at all what one's introspective representations are like. So we are still left with the mystery why tokening *such* an introspective representation would lead one to experience the tomato like *that.*

Once again we are left with a question, not an answer: Why does the tokening of a symbol that is simple in *this* way lead one to have a tomato experience that is simple in *that* way? Spelled out a little more fully, the question reads: Why does introspecting one's tomato perception by tokening a Mentalese lexeme that is semantically simple lead one to have a phenomenologically simple red quale? This is a genuine question. For all I have said, Lycan's augmented HOP theory may still be the true account of the illusion of concrete secondary quality. But if that were so, we would be in the unfortunate position of not being able to understand a deep truth about ourselves.

More basic than these specific worries concerning the question whether the augmented HOP can explain such features as the apparent simplicity and homogeneity of phenomenal features of experience, is the issue whether the theory sheds any light on the question why introspecting perceptual states should feel like anything at all. As far as I can tell, augmenting HOP with the language of thought hypothesis does not put it in a better position to address this quite general version of the having-problem. At best, we have been given a theory that can explain why, *if* there is something it is like to see a tomato at all, the tomato will be experienced as being simply and homogeneously red. This still leaves us with the big question why meta-representing a tomato should feel like anything at all in the first place.

Having voiced this quite general complaint about HOP and the having-problem, I should note that I find Lycan's account of the illusion of concrete secondary quality the single most impressive piece of theorizing to come out of the broad-based effort to understand consciousness in terms of representation. I believe that if our world is a world of relocated$_{(w \text{ or } h)}$ celluloid qualia, then a re-

fined version of Lycan's augmented HOP theory of consciousness has the best shot at explaining why we experience the world in the way we do.

After this excursus on Lycan's ingenious blend of HOP with the language of thought theory, I now return to Armstrong's account of the origin of the illusion of concrete secondary quality.

9.3.5 The Origin of the Illusion

Three facts conspire, Armstrong tells us, to mislead us into accepting the view that we are directly acquainted with simple, homogeneous color qualia. First, there is our tendency to see complex structures as simpler *Gestalts*; second, our grasp of the resemblances and differences between colors makes us think that we grasp the intrinsic natures of the colors themselves; third, we fall prey to the scope fallacy so memorably illustrated by the headless-woman illusion. The intuitive force of these considerations is undeniable. But I think they tend to lose some of their persuasiveness on closer inspection.

9.3.5.1 The Gestalt *Model*

DIRE realism makes the colors into microphysical primary-property complexes. According to the WRT these external properties are the qualia of our experience. The question arises how to account for the discrepancy between the colors as they really are and the colors as they show themselves to us as qualia of our experience. An appeal to the *Gestalt* idea supplied by *Gestalt* Psychology plays a big role in Armstrong's answer: we grasp the relevant microphysical properties in a *Gestalt* fashion that, as it were, 'cloaks' all the complexity in an impenetrable mantle of homogeneity and simplicity.

It is not immediately obvious which of the principles of *Gestalt* psychology to bring to bear on our case. The famous *Gestalt* principles of perceptual organization—good figure, similarity, good continuation, proximity or nearness, common fate, meaningfulness or familiarity, and figure-ground segregation—do not seem readily applicable. The jump from the microphysical surface properties of the tomato to the homogeneous red that suffuses my visual field is markedly different from the sorts of transitions illustrated by the famous drawings with which the *Gestaltists* liked to explain the workings of their principles.

But there is also a deeper and darker side to *Gestalt* psychology. Here is how one author describes the "key emphases" of *gestalt* psychology: "Holistic concepts, not merely as additive sums of the parts, but as emergent phenomena in their own right." (Sternberg 1995: 52) In short: the whole is greater than the sum of its parts. On this model the simple homogeneous quality we experience is understood as a quality that emerges from the underlying structured complex of

primary properties with which the DIRE realist identifies the color. The flavor of this proposal is quite out of keeping with Armstrong's DIRE realism (and out of keeping with everything else that is close to Armstrong's heart!). It looks like the *Gestaltists* may have given him an unwanted gift: a free floating extra property irreducible to the underlying microphysical facts.[129]

9.3.5.2 Resemblances and Intrinsic Natures

I find myself unpersuaded by the second consideration Armstrong marshals in his attempt to explain the origin of the illusion of concrete secondary quality. He argues (see page 237) that our ability to detect resemblance relations among the colors gives rise to our mistaken belief that we are acquainted with simple, homogeneous color properties. For, without this mistaken belief, our ability to detect the resemblances would mystify us.[130] This step of Armstrong's defense of DIRE realism strikes me as implausible. It is implausible because the intrinsic qualities with which we mistakenly take ourselves to be acquainted are not at all the kinds of qualities one would expect, given that Armstrong told the right story about the origin of our belief in them. If what gives rise to the illusion of concrete secondary quality is our failure to perceive the true causes of perceived color resemblances, together with our need to make sense of them, then one would expect the intrinsic qualities with which we feign ourselves acquainted to be such as to explain, or illuminate, or make sense of the perceived resemblances between colors. But this is not so. That the illusory intrinsic qualities resemble each other in the ways they do seems brute and inexplicable to us. That phenomenal green resembles phenomenal blue more than it resembles phenomenal red is just a brute phenomenological fact. It seems, therefore, that if a need to make sense of perceived resemblance is at the root of the illusion of intrinsic quality, then the illusion did not serve its purpose well. For it does not advance our understanding of the problematic fact.

There are other philosophical 'feigning' theories that fare better. Hume—who was much troubled by the bruteness of the resemblance relations between our simple color ideas—proposed a feigning theory of our belief in personal identity. The idea of a simple uninterrupted self is not only not underwritten by any impression but is positively controverted by the variability and interruptedness of all our impressions. How, then, do we manage to mistake an ever changing bundle of distinct but resembling perceptions for an unchanging self? Hume holds that we can perform this feat by feigning acquaintance with a self that introduces the needed unity into the diversity. He writes:

> In order to justify to ourselves this absurdity, we often feign some new and unintelligible principle, that connects the objects together, and prevents their interruption or variation. Thus we feign the continued existence of the per-

> ceptions of our senses, to remove the interruption; and run into the notion of a *soul,* and *self,* and *substance,* to disguise the variation. (Hume 1973: 254)

The truth of Hume's story does not concern us here. But even if it is false it is explanatory: the feigning he ascribes to us helps to explain how we manage to believe in personal identity; the unchanging *self* is the right kind of thing to give us personal identity through time. Armstrong's story, on the other hand, lacks this virtue. The feigned acquaintance with intrinsic properties of colors—simplicity and homogeneity—does not help explain the perceived similarities between them. For this reason, Armstrong's speculative account of the origin of the belief that we are directly acquainted with the intrinsic natures of secondary qualities seems implausible.

9.3.5.3 The Headless-Woman Illusion

The analogy that Armstrong sees between the headless-woman illusion and the illusion of concrete secondary quality (see page 237) has an undeniable appeal. It is illuminating, and it does point to an invalid pattern of inference involving an illicit shift in the scope of the negation. As in the case of the headless-woman illusion, we move from the unawareness of the existence of x to the awareness of the nonexistence of x. We don't see any structure in the red quale and invalidly conclude that we see its structurelessness. The invalidity of the transition is beyond question. Hence, an appeal to the impression of simplicity and homogeneity cannot be used to conclusively establish the simplicity and homogeneity of phenomenal color. Armstrong clearly wins this round. But from the fact that an argument is not valid it does not follow that an argument has no value. Much of our reasoning is invalid. Some of it is good, some bad. The question I want to address is whether this particular piece of reasoning is a valuable piece of invalid reasoning. The conclusion I will reach is disappointingly dull. The foes of DIRE realism cannot hope to use this consideration to persuade their opponents. And the DIRE realists cannot show that their critics are being unreasonable in appealing to this consideration. So we are left with a stalemate of sorts. These reflections close on a more controversial note. I suggest that the 'illusion of concrete secondary quality' is not an illusion after all. Consequently it cannot be profitably compared with the headless-woman illusion.

9.3.5.4 Should We Trust the Impression of Simplicity?

Notwithstanding its invalidity, the pattern of reasoning that takes one from the unawareness of x to the awareness of the absence of x is used frequently, successfully, and with little difficulty. We know when it is appropriate to proceed in this manner. We know that not seeing the Mercury Tracer in the garage is a perfectly good reason for believing that the car is not there. And we know that not

seeing a trace of mercury in the garage is a terrible reason for thinking that there aren't some atoms of this element in the garage. Reasoning in this manner is useful in cases that resemble the Mercury Tracer example and useless in cases that resemble the trace of mercury example. The difference between the two cases has to do with differences in the way the garage would look given the presence or absence of the car and the element respectively. For in assessing the trustworthyness of your inference you have to determine why you did not see what you were looking for. Was it because (i) it was not there? or because (ii) it is the sort of thing you would not have seen even if it had been there? Reason will have been served only if you can determine that you arrived at your conclusion via route (i). To convince yourself that (i) was the case, you will have to know (a) what it would look like if the object were there, (b) what it would look like if the object were not there, (c) that you can discriminate between these two scenarios, and (d) that the scenario you saw is the one that goes with the absence of the object. Conditions (a)-(d) are easily satisfied in the case of the Mercury Tracer. So you are justified in assuming that you arrived at your conclusion because (i) was the case, and you can reasonably assume that your conclusion is true. But you do not get this result when the example is switched from the Mercury Tracer to traces of mercury. Ordinarily you do not satisfy conditions (a)-(d) for mercury traces in your garage. Lacking a justification for the belief that (i) rather than (ii) was the case, it would be unreasonable for you to assume that your conclusion is true.

How, given this background, ought one to judge the reasonableness of the inference from the simple appearance of the red quale to the simplicity of the red quale? It seems that both parties in this dispute would beg the question against each other, if they were to rely on the considerations set out in the preceding paragraph to assess the reasonableness of the problematic piece of reasoning about qualia.

Let's begin by looking at the reflections the anti-reductionist qualophile has to engage in, if she is to convince herself of the reasonableness of her inference. She will have to determine that she arrived at her conclusion that qualia are simple and lack structure via route (i). To do so she will have to know (a) what it would look like if qualia had the complexity and structure posited by the DIRE realist, (b) what it would look like if qualia lacked this complexity and structure, (c) that she can discriminate between these two ways things might look, and (d) that the way things looked to her was the way things would look if the qualia lacked complexity and structure.

In assuming that she satisfies (a)-(d) she is assuming that the DIRE realist's complex, structured qualia would look different than the simple, homogeneous qualia she believes in. Presumably, she believes that, while her qualia would

look simple and homogeneous, the qualia of the DIRE realist just couldn't look that way. It may not be easy to spell out just what they would look like. But, in any case, they would not look simple and homogeneous. Having assumed this striking difference between the two qualia scenarios, she feels confident that not noticing a complex pattern of hatches on the tomato's surface puts her in a position to reasonably conclude that the qualia involved in this experience lack complexity and structure.

Reasoning in this manner, the anti-reductionist qualophile begs the question against DIRE realism. DIRE realism holds that the complex, structured qualia borne by external objects might, and as a matter of fact do, look simple and homogeneous to us. Since her case against DIRE realism turns on denying this claim, her case against DIRE realism is revealed as a *petitio principii*. Unable to establish the reasonableness of her inference to the simplicity and homogeneity of qualia in a non question begging manner, the anti-reductionist qualophile cannot use this consideration in her attack on DIRE realism. But her argument may be of value nonetheless. If the DIRE realist tries to establish the stronger conclusion that her argument is unreasonable, he too gets caught in the trap into which his opponent fell.

So let's turn to look at the way in which the DIRE realist might try to establish that the inference of the anti-reductionist qualophile is unreasonable. He will have to show that she arrived at her conclusion via route (ii). To do that he will have to show that she does not satisfy conditions (a)-(d). The DIRE realist will target condition (c) and argue that his opponent does not satisfy it: she is in no position to discriminate between the simple and the complex qualia scenario. In making this claim, he is assuming that the complex and structured qualia of DIRE realism would present themselves to her in precisely the way that she wanted to reserve for her simple and homogeneous qualia. As the DIRE realist sees it, her confidence in the conclusion that qualia are simple and homogeneous is therefore misplaced. She has no justification for the belief that her failing to notice structure in her qualia is due to its absence. He contends that the complexity and structure of qualia is precisely the sort of thing the presence of which would go unnoticed even if it were there. Hence she is unjustified in assuming that she arrived at her conclusion via (i) rather than (ii). She would, therefore, be acting unreasonably in assuming that her conclusion is true.

But reasoning in this way the reductionist realist assumes what is to be proven. He assumes that the structured primary-property complexes would appear simple and homogeneous to us. This assumption is precisely what his opponent wants to call into doubt. In her eyes the fact that his attack on her argument turns on the problematic assumption will rob it of all its force.

Thus, this discussion ends in a standoff not apt to please anybody. The anti-reductionist qualophile cannot hope to use her argument effectively against DIRE realism. And the DIRE realist cannot hope to convince his opponent that her argument is unreasonable.

9.3.5.5 'Illusion' of Concrete Secondary Quality?

Finally I turn to the second and more controversial argument advertised above. On reflection the analogy between the 'illusion of concrete secondary quality' and the headless woman illusion (or any of the common visual illusions, for that matter) tends to become quite elusive. I suggest that the reason for this disanalogy lies in the fact that the mistake we allegedly make in perceiving colors is quite unlike the mistakes we make in other visual illusions.

This consideration turns on a somewhat controversial assumption. It is this: if you claim that your perception of x contains an element of illusion, then you have to be able to say what it would be like to perceive x without this admixture of illusion. In the case of the headless-woman illusion, for example, what you would have to see is that there is a woman with a black cloth over her head on the stage. I suspect some lesser illusionists have had the misfortune of providing their audience with just this sort of unexciting experience. Other illusions are less trivial—they are impenetrable in the sense that we are doomed to fall for them. Take, for example, the famous Muller-Lyer illusion. The two arrows look to be of unequal length no matter how often you have determined their equality through measurement. Nevertheless, it is easy to say what it would be like to see the Muller-Lyer arrows without the illusory element. It would be to see the two lines as being of equal length.

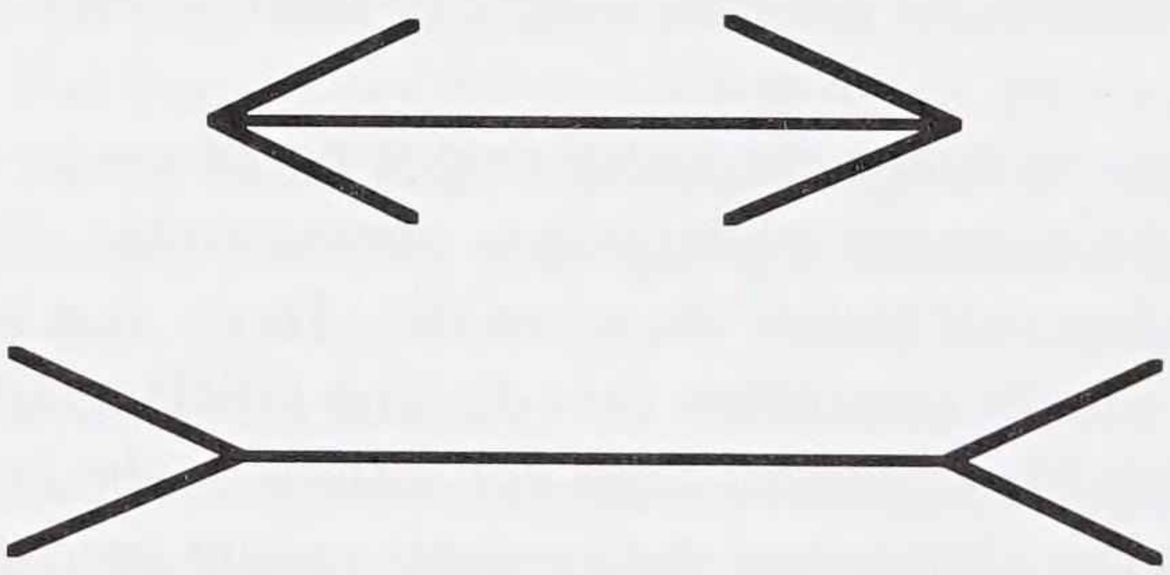


Figure 2. The Muller-Lyer Illusion

So the question that the DIRE realist has to answer is this: What would it be like to see the colors without the illusory component? I do not think that this question can be answered. Reasons for this opinion will be presented in the following section. For now I merely want to point out that this observation, if it holds up, undermines the analogy between the illusion of secondary quality and the headless-woman illusion, thus weakening Armstrong's account of the origin of the illusion of concrete secondary quality. What it suggests is that the so-

called illusion of secondary quality is wrongly classified as an illusion. The mistake we make in apprehending the colors is unlike the mistakes we make in other visual illusions. The very claim that we do make a mistake in grasping the colors in the way we do is rendered questionable by our inability to say what an unmistaken grasp of the colors would be like.[131]

It is worth pointing out that the considerations of this section are easily connected with the argument in the preceding section. The assessment of the reasonableness of the phenomenologically based conclusion that qualia are simple and homogeneous turned on a comparison of what things would look like if qualia were simple with what things would look like if qualia were complex and structured. To make the comparison, a knowledge of both scenarios is required. Assuming that what is required is a knowledge of the scenarios that is not shot through with illusory components, we now have reason to think that nobody is in a position to make the requisite comparison. The defense as well as the critique of the rationality of the problematic inference were based on the assumption that such a comparison could be made. Therefore, both of these arguments fall through. If anything, this result seems to solidify the standoff with which the previous section closed.

9.3.6 The Depth of the Illusion

The appeal to illusion plays a crucial role in Armstrong's version of DIRE realism. One cannot hope to understand the theory without a clear grasp of the nature and the extent or depth of the illusion involved. I have already presented some problems that make it difficult to determine the precise character of the illusion of concrete secondary quality. I suspect that the appeal to illusion remains unintelligible within the DIRE-realist framework. Setting aside these worries about the nature of the illusion, we can ask how deep the illusion runs. I will argue that the DIRE-realist account of color entails that our experience of color is totally illusory. It is therefore misleading to speak of the 'element' of illusion involved in our experience of color. Instead the DIRE realists should speak of the illusion of color. This proliferation of illusion has serious consequences for the WRT (when coupled with a DIRE realism about color). All the problems the WRT faced in the domain of nonveridical experience now threaten to arise within the domain of 'veridical' experience.

The question I want to ask is this: Assume that I am looking at a red tomato and it looks red to me. What exactly is the illusory component of this experience of mine? The search for an answer to this question is guided by the maxim that special attention must be paid to the phenomenal facts involved. The phenomenal red is, after all, the problematic property that needs to be accounted for. We must

make sure that the 'phenomenality' of the phenomenal red does not disappear in the complexities of DIRE realism.

This maxim cannot be pushed aside as something that is gratuitously grafted upon DIRE realism. A condition on the adequacy of the DIRE-realist analysis of phenomenal red is that, after all is said and done, we are still left with *phenomenal* red. For reductive accounts are conservative—they are supposed to conserve rather than eliminate the phenomenon they account for. Applied to the example of my viewing a red tomato, this requirement on a satisfactory reductive analysis of phenomenal red comes to this. Assume that the DIRE realist is given the magical power to make the tomato I see have the property with which he identifies phenomenal red. Then a minimal condition on the correctness of the DIRE realist's analysis is that the tomato does not suddenly look differently to me at this miraculous moment. This is the sense in which the analysandum (if it happens to be phenomenal red) and the analysans must both be phenomenal reds. Perhaps this requirement can be summed up by saying that the DIRE-realist analysis of phenomenal red has to save the phenomena.[132]

9.3.6.1 First Interpretation: Total Illusion

Perhaps we can isolate the element of illusion in my experience of the red tomato by asking the following question: What would it be like for me to experience the phenomenal redness of the tomato without any illusory admixture? Subtraction of this 'disillusioned' experience of phenomenal redness from the ordinary experience of phenomenal redness will yield the illusory component as the remainder. Recalling Armstrong's claim that "secondary properties are primary properties not apprehended as such," a disillusioned apprehension of the tomato's phenomenal redness should amount to my apprehending the complex of primary properties that is its phenomenal redness *as such,* i.e., as a complex of the kinds of properties recognized by physics. But when I try to imagine what it would be like for me to see the red tomato in this disillusioned manner the mind boggles. Am I supposed to see a conglomerate of primary properties *and nothing else*? What would the surface of the tomato look like if it were experienced as having 'bare' primary properties? I have no idea what to imagine. More than that, I think that this notion of disillusioned experience of phenomenal red is inconceivable because incoherent. Berkeley was right in maintaining that the idea of a primary property abstracted from all other sensible qualities "implies a repugnancy in its conception." (Berkeley 1985: 29) Consequently, we cannot 'frame the idea' of phenomenal red as it really is, uncontaminated by illusory elements. There is no such thing as a disillusioned experience of phenomenal redness. The experience of phenomenal redness is *all* illusion. That is, if I subtract the illusory component out of my current experience of the tomato's redness, the phenome-

non changes completely: the very arresting and striking phenomenal red is replaced by an absence of anything phenomenal. The 'phenomenality' of the phenomenal red disappears without remainder. The illusory component of my experience of phenomenal red accounts for the whole of my experience. There is nothing non-illusory about it. By declaring the phenomenal aspect of phenomenal red to be illusory, DIRE realism turns into a peculiar form of eliminativism. Instead of telling us that phenomenal red exists and what it really consists in, it shunts the phenomenal dimension into the realm of the merely illusory which, by its very nature, need not be accounted for. If this is the correct interpretation of the DIRE-realist analysis of phenomenal redness, then I would regard DIRE realism about color as a failure.

9.3.6.2 Second Interpretation: Partial Illusion

Here is a different model with which to interpret the notion of disillusioned experience. Consider the Benham disk. When the back and white disk rotates slowly one clearly sees a number of differently colored concentric rings of color on its surface. One might want to explore the following analogy. Looking at the spinning Benham disk is analogous to the normal experience of the phenomenally red tomato. Looking at the static Benham disk is analogous to a disillusioned experience of the phenomenally red tomato. Stopping the Benham disk reveals the complex of achromatic surface features that is the spinning disk's phenomenal redness. When looking at the static disk, I apprehend the complex of achromatic properties that is the spinning disk's redness *as such,* i.e., I apprehend the disk's redness as the black and white pattern it is. If we could 'stop the tomato' in a manner analogous to which we can stop the rotating Benham disk, we might be able to apprehend the complex of achromatic properties that is its phenomenal redness *as such,* i.e., as the black and white pattern it is.

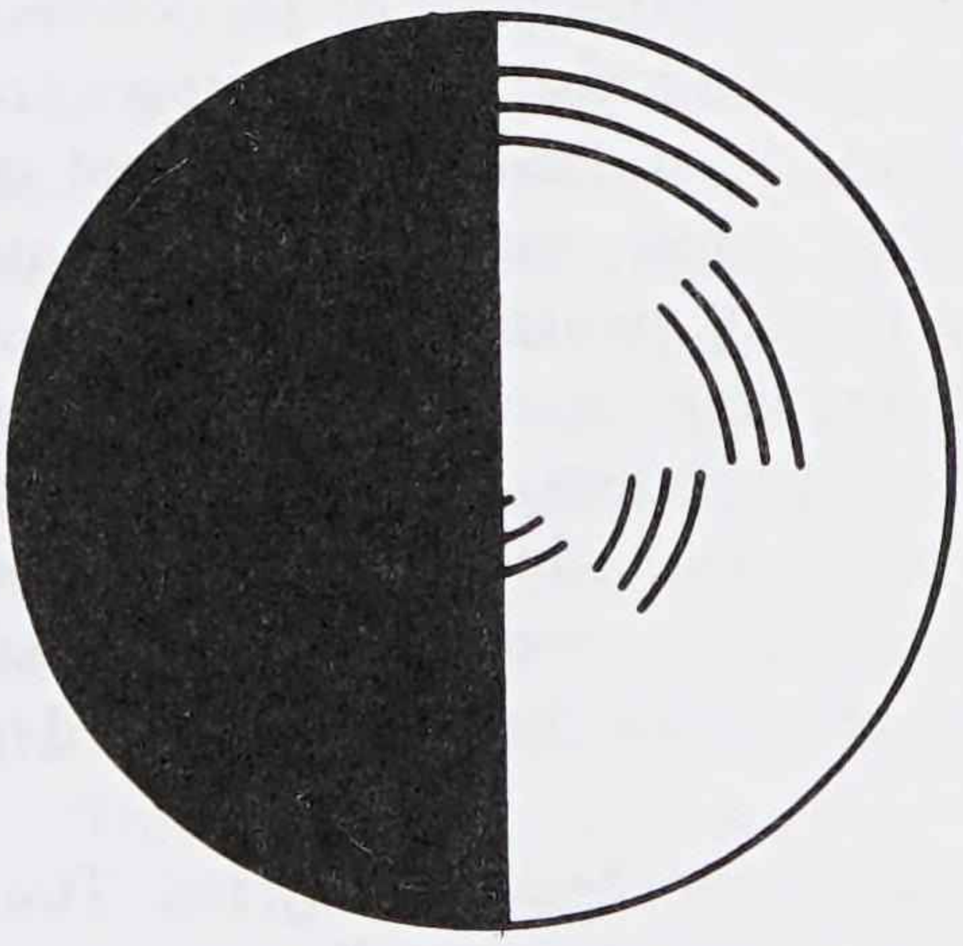


Figure 3. The Benham Disk

What does most of the work in this analogy is the switch from the opposition between primary and secondary qualities to the opposition between achromatic and chromatic properties. Substituting the latter opposition for the former is a seductive and misleading move. The difficulty we experience in trying to imagine a world devoid of secondary qualities is equaled by the ease with which we

can envision an achromatic world. But the achromatic world is no closer to the world of the physicist than the chromatic world. We tend to overlook this because we are prone to forget that achromatic colors are colors. Our tendency to identify color with chromatic colors lets us think that the achromatic world is the physicist's world of primary properties. But from the physicist's point of view, achromatic and chromatic colors are equally problematic features of the world. Thus, an account of color in general in terms of achromatically colored features of objects could not be the whole story. Nevertheless, it would be a remarkable reductive achievement. Can we use it to gain a more satisfactory interpretation of DIRE realism?

One obvious advantage of this interpretation of the nature of disillusioned experience is that it does not succumb to Berkeley's criticism. For this model no longer requires us to experience complexes of 'bare' primary properties. But the failure to be absurd is not sufficient to recommend a doctrine.

To get a better feeling for the nature and extent of the illusion that this interpretation of DIRE realism postulates in the ordinary experience of phenomenal color, we will repeat the experiment of imagining what a disillusioned experience would be like. If I managed to 'stop the tomato,' its surface would present an intricate pattern of microscopic structures in black, white, and grey. There is no incoherence in this. What we imagine seeing is very similar to what one sees upon looking at colored objects through the microscope. But if the disillusioned experience of the tomato is the experience of a grey pattern, we have to ask the question: Where did the red go? Subtraction of the disillusioned experience (the grey pattern) from the ordinary experience (the simple continuous phenomenal red) does seem to leave an uncomfortably large remainder: the phenomenal red in its entirety.

A tomato with a highly structured grey surface is not phenomenally red. Thus it seems that the red that we see when we look at the tomato is an illusion. Again we have to say that the illusion is complete: we experience simple and unanalyzable phenomenal redness where there is nothing of the kind. But a theory that declares phenomenal redness to be illusory is not a reductive theory of phenomenal redness. Hence, the second interpretation of the DIRE-realist analysis of phenomenal redness is no more satisfactory than the first. Instead of explaining how a physical object can be phenomenally red, it eliminates phenomenal redness by making it wholly illusory.

9.3.6.3 Third Interpretation: Minimal Illusion

The first two attempts to understand the illusory component of DIRE realism went wrong by making the illusory component too pervasive. The illusory component did not stay confined to some element of phenomenal redness. Phenome-

nal redness as such ended up as an entirely illusory property. What is needed is an interpretation of the illusion that properly restricts the extent of the illusory element.

Perhaps we can achieve a better understanding of the nature of the illusion by following up on Armstrong's remarks on this issue. As we have seen, he suggests that we are "aware of [the secondary properties] in a unified, *Gestalt,* manner, a manner which fails to reveal the primary nature of these properties." (Armstrong 1984: 178; see also Armstrong (1977: 30)) He further elaborates this claim by saying that perception "fails to make us aware of any great complexity in the phenomenal qualities, and further fails to make us aware of an identity of these qualities with complex physical properties." (Armstrong 1973: 50) Thus we suffer a twofold illusion when we are aware of a secondary quality: "the illusion that it is not a primary quality" and "the illusion that it is not a structured property." (Armstrong 1984: 179)

Let us focus on the second kind of illusion, the illusion that "phenomenologically, the secondary qualities lack structure." (Armstrong 1984: 180) How would things change if we could rid ourselves of this illusion? What would it be like to experience phenomenal red with the appropriate structure? Armstrong's appeal to *Gestalt* phenomena may incline one to think of the kinds of illusions we can experience with mosaics or pointillist paintings. In both cases we can be led to experience a structured aggregate of colored dots as a homogeneously colored expanse. By approaching closer to the painting or mosaic, we can get a disillusioned view: the homogeneously colored expanse dissolves into a multitude of colored specks. Thus the illusion of structurelessness is easily dispelled in these cases. Can this serve as a model for the illusion of simplicity?

This model does suggest a much less radical interpretation of the illusory element of DIRE realism. The illusion consists in a filling in of gaps according to some *Gestalt* principle of closure. Thus we suffer the illusion of homogeneity where there really is a discontinuous multiplicity. But the multiplicity involved is not one of 'bare' primary properties or of achromatic properties; rather it is a multiplicity of phenomenally colored particulars. There is, then, a sense in which this version of DIRE realism is not guilty of eliminating the phenomenon of which it purports to present a reductive analysis.

But to the extent that this interpretation succeeds in saving the phenomena it fails to be genuinely reductive. The phenomena are saved because the analysis is carried out in the very terms that are to be analyzed: the notion of a homogeneously colored expanse is analyzed in terms of the notion of a collection of homogeneously colored dots. This strategy pushes the problem one step back but does not solve it. The attempt to complete this kind of analysis by applying it repeatedly to ever smaller colored expanses does not promise success, for it is doubt-

ful that the regress could be terminated in a satisfactory manner. The regress could come to an end in case there existed phenomenally red elementary particles in terms of whose redness the redness of larger objects is to be explained. But this halting point is unsatisfactory from the DIRE-realist perspective because the analysans still contains the notion of phenomenal redness. If, on the other hand, the phenomenal redness of suitable small expanses is understood in terms of complexes of bare primary properties or in terms of complex achromatic qualities, we run into the problems that were surveyed above.

9.3.7 DIRE Realism and Illusory Colors

I find myself unable to provide a satisfactory interpretation of the nature of the illusion that, according to DIRE realism, is involved in every experience of any phenomenal color property.[133] Phenomenal colors as we experience them seem to be illusions through and through. The 'element of illusion' involved in seeing the phenomenal red of the tomato is not merely a minor factor causing a slight distortion in our manner of experiencing the tomato—the illusion is complete insofar as phenomenal color is involved. The color realist is forced to banish the specifically phenomenal aspect of phenomenal red into the realm of illusion. While this move may smooth the path to a reductive account of phenomenal red—or rather, those features of phenomenal red that are not phenomenal (whatever they may be)—it does leave us with the problem of accounting for the banished illusory colors. To say that they are unreal because illusory and are, therefore, not in need of further explanation will not do. Illusory colors are colors and as such they present a problem for the would be materialist.

By declaring phenomenal colors illusory (in the sense of not being literally ascribable to the external physical objects we see) one creates a problem rather than solving one. For at this point the return of the ill-fated merely intentional objects seems inevitable. If the real tomato is not allowed to bear the phenomenal redness I experience it as having, then the merely intentional tomato will have to stand in as a surrogate bearer of the experienced phenomenal property. And the difficulties involved in countenancing phenomenally colored merely intentional objects have been canvassed *ad nauseam* above.

Now we have come full circle. The discussion of the DIRE-realist interpretation of the relocation thesis started out with a consideration of the problems that nonveridical experience poses for this theory. And these problems were formidable. But since the theory's ability to handle cases of veridical experience seemed much more relevant for the assessment of the success of the theory, we then turned to the question how the theory fares in the domain of veridical experience. At first sight the theory appears to be much better suited to handle veridi-

cal experiences. But the close scrutiny of the case of veridical experiences of phenomenal color has proven this impression wrong. The DIRE realist is forced to use the very same problematic notions in his analysis of veridical and non-veridical experiences of phenomenal colors. Thus, DIRE realism is in no better shape in the domain of veridical experience. The problems afflicting this theory are not a minor anomaly arising from the odd cases of nonveridical experience. All experience of secondary qualities, be it veridical or nonveridical, raises the same profound problems for the DIRE-realist interpretation of the relocation thesis.

One final comment. We are now also in a better position to appreciate the force of a point alluded to above on page 218. The introduction of merely intentional qualia-bearing objects is useful only to the extent that the designated qualia-bearers can do their job. Since the philosophers of the DIRE-realist persuasion tend to be tough-minded materialists, their otherworldly intentional objects are regular physical objects. Thus, the problems posed by the argument from atomism apply all over again. Merely intentional physical objects are no better positioned to exemplify qualia than the physical objects of this world.[134]

9.4 Summary

In this chapter we took on the central claim of the relocation theses. On the grounds that the Mind cannot house qualia the relocationist holds that qualia must be moved out of the Mind. In its wholehearted version relocationism pushes the qualia all the way out onto the intentional object. The halfhearted version of the doctrine is content to move the qualia into the brain. The arguments in this chapter apply to both versions of the relocation thesis. But for simplicity's sake the wholehearted version of the doctrine was our focus.

Two traditional objections to the relocation move were raised. First, the argument from illusion: there are no external objects to bear the qualia of nonveridical experience. Second, the argument from atomism: qualia are not the kind of property a physical object can bear.

In response to the problems raised by nonveridical experience, the relocationist$_w$ claimed that (at least in these problematic cases) qualia are borne by merely intentional objects. But intentional objects are not a good home for qualia. On one construal (inspired by the early Brentano) intentional objects are mental and therefore undermine the relocationist's$_w$ project of rendering qualia nonmental. Meinongian intentional objects neither have a home (they are, in his famous phrase, homeless objects) nor do they make a good home for qualia. Mere possibilia—whether of a Lewisian or Ersatzist variety—fare no better. The

exemplification-problem and the causation-problem stand in the way of identifying intentional objects with mere possibilia. The relocationist's$_w$ ends cannot be served by an intentional object that is a mere possibile. Though the argument focused on the WRT the problems of nonveridical experience are not limited to it. The HRT confronts parallel questions when a higher-order representation is directed at a lower-order mental state that does not exist. It seems, then, that nonveridical experience poses a formidable problem to the relocation theses. But since nonveridical experience only constitutes an anomalous case, not too much weight should be attached to it.

The argument from atomism takes on the relocation theses in the domain of veridical experience. I focus on Armstrong's DIRE-realist attempt to deal with the argument from atomism. Armstrong grants the atomist his conception of matter and revises our notion of a secondary quality so that it fits into the atomistic picture. By claiming that secondary qualities really are complexes of primary properties, Armstrong can embrace the atomistic picture of matter and claim that the phenomenal properties are properties of material objects. To save the phenomena, the DIRE realist has to claim that the experience of secondary qualities contains an element of illusion. This is the claim I criticize in the remainder of the argument.

I argued that the nature of the illusion of concrete secondary quality is problematic. Neither Armstrong nor Lycan provide a phenomenologically adequate account of the nature of the illusion. Turning to the origin of the illusion, I focused primarily on one of Armstrong's ingenious suggestions designed to explain the illusion of concrete secondary quality: the headless-woman illusion. Having granted that Armstrong's observation is of considerable force I ended doubting the analogy between the two illusions. Assessing the depth of the illusion, I argued that DIRE realism turns phenomenal qualities into total illusions. Thus, it seems that the relocationist who attempts a DIRE-realist explanation of veridical color experience can only succeed if he has an adequate account of nonveridical or illusory color experience. But I have already argued that nonveridical experience is more than the relocation theorist can handle. The relocationist's explanation of the veridical and the nonveridical experience of phenomenal properties stand or fall together. I have tried to show that they fall together.

Where does this negative result leave us? It would seem that we have reached something of an impasse. The friends of deMented qualia have argued that the Mind contains no qualia-bearers. I have argued that the nonMental part of the world contains no qualia-bearers. It seems to follow that there are no qualitative phenomena since there exists nothing that can instantiate or exemplify qualia. But this is clearly wrong. In the next chapter I will propose a way of resolving

this impasse. I will argue that the hunt for qualia-bearers is altogether misguided. What is needed is an account of qualitative phenomena that waives the requirement of a qualia-bearer.

CHAPTER TEN

Consciousness: The Having of Qualia

10.1 Learning from Mistakes

To be conscious is to have qualia. How much progress have we made in elucidating this somewhat cryptic slogan? Let's begin by taking stock.

First, What have we learned about the having of qualia? A large part of the discussion of the having-problem was devoted to a critique of introspectionism. In Chapter Four I argued that introspectionism was and still is the predominant account of how we have our qualia. In Chapter Five I tried to explain, and thereby explain away, the undeniable attractiveness of introspectionism. In Chapter Six I argued that we have good reasons for thinking introspectionism false. Chapter Eight broadened the scope of the discussion to include all representation-based accounts of how relocated qualia are had. Relocating qualia (wholeheartedly or halfheartedly) makes the having-question intractable. I concluded that having a relocated quale is not a manner of representing it, no matter what the details of representational activity may be. Since there do not seem to be any other (i.e., non-representation-based) accounts of how relocated qualia are had, we were forced to conclude that we do not know what it is to have a relocated quale.

Second, What have we learned about the nature of qualia? Early on (Section 2.2.2.3) I introduced the neutrality thesis, according to which the notion of a quale is to be understood on a purely phenomenological level without any commitment to the metaphysical status of qualia. But much of the discussion of introspectionism (especially the first appeal to phenomenology in Chapter Five) seemed to be in keeping with the traditional assumption of the inner, Mental (i.e., nonphysical) nature of qualia. The second appeal to phenomenology (Section 7.4.1) shattered the time-honored idea of the phenomenologically revealed innerness or Mentality of qualia. The further observation that there are no Mental qualia-bearers (e.g., sense-data) seemed to clinch the case against the Mentality of qualia. So qualia are *not* Mental properties. DeMenting and relocat-

ing qualia seemed to be the only way out. This expulsion of qualia from the Mind is invariably (but perhaps not necessarily) accompanied with the substitution of celluloid qualia for silver-screen qualia. Both versions of relocationism proved problematic. The intractability of the having-problem for relocated celluloid qualia (Chapter Eight) was already mentioned. And in Chapter Nine I argued that the problems raised by nonveridical experience and by the illusory component of veridical experience show that qualia cannot be relocated. There are no physical objects—not in the head or outside of it, not in this world or in other worlds—fit to function as bearers for qualia. So qualia are *not* physical (or better: not nonMental) properties.

In sum, we have learned nothing positive about either one of our two questions. We do not know what it is to have qualia. And we do not know what the nature of qualia is. It would seem, then, that our understanding of the claim that to be conscious is to have qualia has not been advanced. All we can say is that we are conscious because we are related to we know not what in a manner we know not how. And that is not saying much.

But this does not mean that we have learned nothing of value. By learning what qualia aren't and what it isn't to have them, we have learned a great deal about where to look for qualia and about how to think about what it is to have them. The negative results are not strong enough to single out a unique account of consciousness. But they do suggest a direction for further research. First, we want qualia that are deMented and unrelocated. For we agreed that there are no Mental qualia-bearers and discovered that there are no nonMental qualia-bearers either. The only way to satisfy both of these demands—short of eliminating qualia altogether—is to opt for bearer-less qualia—qualia that need no bearers to be exemplified. Second, we want silver-screen qualia. I have no story to tell about how representing a quale lets you have it. Thus, we have to reject celluloid qualia—qualia that need to be represented in order to be had—and to opt for self-presenting silver-screen qualia—qualia that are had just in case they are instantiated.[135] In this final chapter I will, therefore, attempt to develop an account of consciousness in terms of unborne qualia the having of which consists in their presenting themselves to us.

10.2 Isolating Two Pernicious Assumptions

To ensure that the proposed account of consciousness avoids the mistakes of its rivals, its construction must be informed by the lessons learned from their mistakes. Since one can learn from mistakes only after one has clearly grasped them, I begin by isolating the roots of the problems besetting the rival accounts.

We encountered two basic problems. First, there is the problem of the nature of the objects that bear the qualia of our experience. Second, there is the problem of the nature of the mental acts that let us have our qualia. These two big issues found expression in a host of unanswered questions: What kinds of objects are the qualia-bearers supposed to be? Are they Mental or physical objects? If physical, are they also mental or are they purely physical objects in our environment? Are they actual or merely possible objects? Which of these candidate qualia-bearers can actually exemplify phenomenal properties? In which relation must a qualia-bearing object stand to us so that we can come to have the qualia it bears? How should we think of the mental act that allows us to have qualia once belief, knowledge, introspection, and representation in general have been excluded as possible vehicles for the having of qualia? Since I do not know how to answer these questions, I must attempt to construct the positive account in a way that keeps these questions from arising. Questions that cannot be solved must be dissolved.

Once it is understood why these problems and questions arise in the first place, it becomes clear what we must do to make them go away. The problems arise because of two unquestioned (and often tacit) assumptions. The first concerns the structure of experience: it is assumed that experience is a matter of a subject's relating itself to an object through an act. I will call this 'the assumption of the act/object structure of experience' ('act/object assumption' for short). The second assumption concerns the qualia-bearing object of experience: it is assumed that the relation of qualia-bearer to quale is that of substance to attribute. I will call this 'the assumption of the substance/attribute structure of qualitied objects' ('substance/attribute assumption' for short). If we can develop a conception of qualitied objects that does not require an underlying substance or substratum, then the problem of the intractable nature of qualia-bearing objects of experience vanishes. And if we can come up with a nonrelational analysis of experience with which to replace the act/object model, then the question which mental act lets us have qualia vanishes too. Our goal must be, then, to present an account of consciousness that avoids both problematical assumptions. I will discuss two theories of consciousness that get by without them—adverbialism and monism.

Adverbialism gives up on the act/object structure of experience by, in effect, making the qualitied object into a modification of the subject. Adverbialism is an elegant and radical way of dispensing with the two problematic assumptions. The act/object assumption and the substance/attribute assumption come to the fore only if one wants to answer certain questions. The radicalism of adverbialism consists in the fact that it prevents these kinds of questions from arising in the first place. Since the adverbialist does not countenance qualia (understood as

phenomenal properties), she will not ask how qualia are exemplified and how they are had. Since questions that do not arise need no answer, the two pernicious assumptions have no opportunity to mislead the adverbialist.

Monism is not nearly as radical as adverbialism. Because the monist accepts the existence of qualia (understood as phenomenal properties), he must answer the difficult questions concerning their nature and the manner in which they are had. The attempt to answer these questions without making the two problematical assumptions forces the monist to take a considerable amount of controversial philosophical theory on board. By invoking a bundle theory of phenomenal objects, the monist can avail himself of the conception of a qualitied object that does not commit him to a qualia-bearing substance.[136] And by collapsing act and object into one, he arrives at a notion of a self-presenting percept that need not become the object of a mental act in order to be had. Helping himself to a bundle theory for a second time—a bundle theory of the self—the monist will say that for a percept (with its qualia) to be had is for it to become an element of the bundle that is the experiencing person.

10.3 Adverbialism

Adverbialism is most easily motivated by considering the problems arising from nonveridical experience. Instead of postulating some peculiar kind of substitute object for us to be related to in cases in which we perceive what does not exist, adverbialism presents a construal of experience or sensation that dispenses with the idea that in experience we are related to anything:

> What is ordinarily described as a case of someone's perceiving something that does not exist is really not a case of perception, or of any other kind of consciousness, of an object at all, whether of the object we ordinarily take it to have (a material thing) or of any other object (e.g., a sense-datum). It is really a case of someone's being in a state of consciousness that has a certain nonrelational (whether sortal or qualitative) characteristic. Since a state of consciousness is expressed most naturally with a psychological verb, its nonrelational characteristics would be expressed most naturally with adverbs modifying that verb. Hence the designation...as the *adverbial theory.* (Butchvarov 1980: 263)

The adverbialist's project is to discover a new mode of inherence of the phenomenal in the mind. From this point of view, the debate about the location of qualia-bearing objects is entirely misguided: qualia are not properties of things, be these things mental or physical, actual or merely possible. Thus the adverbialist rejects the presupposition that makes the dispute about the relocation the-

ses possible. The proper function of the terms that were taken to stand for phenomenal properties of intentional objects is to designate kinds of experiencing or sensing. Accordingly, the delirious person must not be described as seeing a nonexistent pink rat, or as seeing a pink ratlike sense-datum, or as seeing a rat-appearance which is pink. Instead, we should say that he senses in a-pink-rat-ly manner. Intentional objects as bearers of phenomenal qualities are thereby rendered superfluous. We are left with only one of the original relata, the experiencer, and his various kinds of sensory states. In this way the adverbial theory makes good the promise of being a nonrelational theory of experience.

So much for a very brief sketch of the central idea of adverbialism. By recategorizing qualia (phenomenal properties) as modes in which mental states or processes go on, the adverbialist appears to have solved (or dissolved) the two problems that have proven so intractable up to now. This makes the adverbialist position very attractive. And it also makes the adverbialist theory of consciousness into the most serious challenger to the sort of theory I endorse. If adverbialism is true, then consciousness is *not* a matter of *having* qualia. For qualia, understood as phenomenal properties of phenomenal objects (no matter what their nature may happen to be), have no place within the adverbialist framework. You cannot have an account of consciousness that is both qualia-based and adverbialist in spirit.[137]

And this is not all that adverbialism has going for it. As will become clear later on, the account of consciousness that I favor comes at a considerable metaphysical cost. In comparison, the adverbialist doctrine may appear reassuringly sane. So much by way of an introduction to adverbialism. I am tempted to say I wish it were true—it would, I think, make for a simpler world. If I could convince myself of the truth of adverbialism I would gladly embrace it.

10.3.1 Problems with Adverbialism

The advertisement for adverbialism makes it look very good indeed. It is no accident that many materialists have opted for it. But I find two problems with the adverbial analysis of experience. The first is somewhat technical. Following the lead of Panayot Butchvarov and William Lycan, I argue that closer inspection of adverbialism reveals that it does not keep its promise. The act/object assumption and the substance/attribute assumption reappear when pressure is applied to the theory. Thus, adverbialism does not live up to its billing of being a genuinely new kind of account of the inherence of the qualitative in the mind. The second objection to adverbialism is intuitive rather than technical—but no less damaging for that. I contend that adverbialism does not pass Searle's 'trying on test.' If our experience were structured as the adverbialists would have it, then our expe-

rience would not be what it is. In short, I doubt the phenomenological adequacy of the adverbial analysis of experience. I am not convinced that either one of these objections is conclusive. But to my mind they are serious enough to motivate a search for a better solution.

10.3.2 The Reappearance of the Two Pernicious Assumptions

Adverbialism tends to get unwieldy when dealing with complex perceptual situations. So let us simplify things by looking at the simplest possible case. Take a situation in which, as one would ordinarily say, I sense something red and which the adverbialist would describe by saying that I am sensing redly. How is the sentence 'I sense redly' to be understood? More particularly, how does the word 'redly' function in this sentence?

I will approach this question in a somewhat indirect manner. Criticizing Brentano's later, reistic, account of intentional objects, Roderick Chisholm once raised a series of simple but insidious questions about Brentano's analysis of thought. Some minor revisions in the wording of these questions allow us to address them to the adverbialist theory of sensing. I hope to show that these questions loose nothing of their critical potential in this process of 'translation.' The prospect of using Chisholm's critique of Brentano's reistic account of intentional objects against the adverbial theory of experience is particularly enticing because Chisholm is one of the leading advocates of adverbialism.[138]

Addressing the later Brentano's theory of thought, Chisholm raises the following question: How is the term 'unicorn' used in the sentence 'John is thinking about a unicorn'?

> We may say, as Brentano suggests, that in "John is thinking about a unicorn" the word "unicorn" is being used to contribute to the description of John. But *how* does it contribute to the description of John? We are *not* saying, obviously, that John is a unicorn. We are saying that John is *thinking* about a unicorn, and so might be tempted to say that the word "unicorn" is being used to describe John's thought. But *how* does the word "unicorn" contribute to the description of John's thought? We are not saying, obviously, that John's thought is a unicorn. We are saying—again obviously—that the *object* of John's thought is a unicorn. But, Brentano tells us, statements ostensibly about the object of John's thought are actually statements about John. And so we have completed a kind of circle. For now we can ask, once again: what does this use of "unicorn" tell us about John? (Chisholm 1970: 145)

These are hard questions to answer. In the paper from which I quote Chisholm uses them to bolster Brentano's earlier ontological interpretation of intentional inexistence. In a different paper Chisholm (1982: 63) uses these same consid-

erations to support a Meinongian construal of intentional objects. I believe that Chisholm's remarks about the role of the word 'unicorn' in the quoted passage carry over, more or less directly, to the manner in which the adverbialist would have us understand the word 'red' as it occurs in 'John senses something red.' If the analogy pans out we may take it to support the act/object analysis of sensation. Here are the Chisholmian questions that we may want to address to the adverbialist:

> We may say, as the adverbialist suggests, that in "John senses something red" [1] the word "red" is being used to contribute to the description of John. But *how* does it contribute to the description of John? [2] We are *not* saying, obviously, that John is red. We are saying that John is *sensing* something red, and so might be tempted to say that [3] the word "red" is being used to describe John's sensing. But *how* does the word "red" contribute to the description of John's sensing? We are not saying, obviously, that [4] John's sensing is red. [5] We are saying—again obviously—that the *object* of John's sensing is red. But, the adverbialist tells us, [6] statements ostensibly about the object of John's sensing are actually statements about John. And so we have completed a kind of circle. For now we can ask again: what does this use of "red" tell us about John?

Do these Chisholmian claims and questions adequately represent adverbialism; and, if so, do they embarrass the adverbialist? Let us begin with [5]. The adverbialist has to agree that this is what seems obvious before the advent of adverbialism. That the term 'red' in 'John senses something red' serves to specify the color of John's intentional object is obvious to all but those who have already embraced adverbialism. [5] only states what the adverbialist knows. It causes him no embarrassment.

I turn to sentences [1] and [6]. Is it true that the adverbialist holds that the terms which we mistakenly took to stand for phenomenal properties of John's intentional object are really about John in the sense of contributing to John's description? The adverbialist, as we have characterized him, does not explicitly say so. Nevertheless, the adverbialist is committed to this claim. By making properties of intentional objects into properties of sensory states of the experiencing subject, the adverbialist does, in effect, acknowledge that the term 'red' in 'John senses something red' serves to describe John rather than the object he senses. In describing the way in which John senses as being of the sort red, I contribute to the description of John. I believe that we do the adverbialist no injustice if we say that [1] and [6] are true of his position. But, again, these two claims need not cause the adverbialist any embarrassment. Moreover, [2] and [3] are unproblematic. For John to be red is one thing; for him to sense redly is another. No adverbialist would want to deny that. And the claim that 'red' is used to de-

scribe John's sensing is unspecific enough to be acceptable to every brand of adverbialist.

This leaves us with [4]. [4] does adequately describe the adverbialist position and it should embarrass the adverbialist. For on its most straightforward reading [4] is false, and obviously so.[139] To say that John's sensing is red is no better than to say that John's thought is a unicorn. Sensings or experiencings are neither red nor are they round, triangular, stinking, noisy, hard, or fruity. If the oddness of the notion of a triangular state of consciousness were not enough to convince the adverbialist of the falsity of sentences such as [4], the following consideration might persuade him. 'John's sensing is red' entails 'something is red.' Hence, if John's sensing redly were just a matter of John's sensing being red, then going adverbial would not have the desired result. For then 'John senses redly' would entail 'There exists something that is red,' which is the very entailment that going adverbial was supposed to block.

It seems, then, that Tye's suspicion has been borne out: the Chisholmian questions have cleared away the verbal smoke screen behind which the adverbialists hid the phenomenal objects with their qualia.

But the adverbialist might reply that [4] admits of an interpretation that is more charitable to his cause. The adverbialist Chisholm, for example, has argued that the 'red' in 'red sensing' is not the 'red' of 'red tomato.' While the latter ascribes a phenomenal property to an object, the former tells us of what *kind* or *sort* the conscious state is. And sometimes he puts this point by saying that the 'red' in 'red sensing' tells us something about the way or manner in which the sensing occurs. In the following passage Chisholm establishes this particular use of adjectives:

> Frequently the word "blue" is used to designate a kind or species of appearing; a thing which may, or may not, have the secondary quality *blue* is then said to appear or to look blue. Many other adjectives are used, similarly, to designate...a way of appearing, or of sensing. (Chisholm 1957: 127)

And the following passage makes it abundantly clear that this is the use to which adjectives are put in the adverbial theory of experience:

> For in saying "He is appeared to white," or "He senses whitely," we are not committed to saying that there *is* a thing—an appearance—of which the word "white," in its sensible use, designates a property. We are saying, rather, that there is a certain state or process—that of being appeared to, or sensing, or experiencing—and we are using the adjective "white," or the adverb "whitely," to describe more specifically the way in which that process occurs. (Chisholm 1966: 96)

What we have here is a clear-cut rejection of the above interpretation of [4] and a new proposal of how to understand [4] correctly. Accordingly, the controversial sentence [4]—'John's sensing is red'—does not say that John is undergoing a process that is red in the way in which a tomato is red. It says that John is sensing in a particular way or, in other words, that his sensing is of the species *red*. And the adverbialist maintains that the fact that John is having a sensation that belongs to the species *red* neither commits one to such odd notions as a red, triangular, stinking state of consciousness nor does it entail that there exists a phenomenally red thing anywhere. Thus, there no longer is a need to postulate an intentional object as bearer of the phenomenal redness. By relying on the new interpretation of the problematic sentence [4], the adverbialist can avoid the unwelcome consequences I wanted to pin on him. The adverbialist has answered the Chisholmian questions without adverting to the phenomenal objects and the qualia they bear. If adverbialism does conceal such objects, more searching questions are needed to uncover them.

The adverbialist gambit leaves me dissatisfied. Is the talk about certain species or manners of sensing more than thinly veiled talk about the sensing of things that are certain ways? In other words, Is there a way of understanding such talk that does not boil down to understanding it as talk about the sensing of things that have certain phenomenal properties? The impression that there may not be a distinctive way of understanding the 'redly' in 'senses redly' is reinforced, if we look at Chisholm's attempts to elucidate such uses of words like 'red(ly)' or 'white(ly).' He tells us that "the word "white" tells us something about the *way* in which the object appears, just as "slowly" may tell us something about the way in which an object moves." (Chisholm 1966: 96) But this analogy does not help improve our understanding of the problematic 'redly' in 'senses redly.' I can do no better than endorse what Butchvarov has said on this matter: "We mean by "He runs slowly" exactly what we mean by "His run(ning) is slow," and if we were told that the former should be so understood that it is not equivalent to the latter, we would avow that then we have no idea of what it means." (Butchvarov 1980: 265) Thus we are faced with a dilemma: either the phrase 'senses redly' remains incomprehensible or we have to interpret it in a manner that threatens to undercut the adverbialist gambit. Since the first horn is unacceptable the adverbialist has to confront the second:

> But we cannot understand "I am appeared whitely to" as meaning the same as "My (present) being appeared to is white," since the latter is either necessarily false or senseless. And if it made sense and also were true, then the curious result would follow, that after all "I am appeared whitely to" does entail that something is white, though not that a material thing or a sense-datum is white, but that a state of consciousness is white! (Butchvarov 1980: 265)

This, then, is the reason for believing that, if we can understand the adverbialist's formulations at all, we can only understand them as attributing phenomenal properties to states of consciousness. Thus, Butchvarov's critique has led us right back to the original objection to the adverbialist theory of experience. The claim that the adverbs in the adverbialist formulas designate species of sensing rather than phenomenal properties of sensings has not managed to displace the original uncharitable interpretation of the problematic sentence [4]. Pending a fuller account of the functioning of the adverbs in adverbialism, the adverbialist theory of experience is incapable of answering the Chisholmian questions in a satisfactory manner.

Inasmuch as we can understand the adverbialist at all, it seems that qualia, understood as phenomenal properties of phenomenal objects, are back again. Thus the subject/attribute model—and all the questions to which it gives rise—is back. The restoration of qualia to their legitimate status as phenomenal properties also brings back the having-problem. The reintroduction of the qualia-bearing phenomenal object destroys the immediacy that characterized the adverbialist account. Instead of a single subject sensing in a certain manner, we now have a subject and a separate object that exemplifies the quale. Inevitably the question will arise how the former manages to *have* the latter in such a way that it thereby becomes like something to be the former. That the having-problem now poses itself in this way shows that the act/object assumption is back as well. The search for the mental act that lets us have our qualia is on again and the adverbialist (dis)solution of the having-problem dissolves.

10.3.3 The Phenomenal Adequacy of Adverbialism

Is the adverbial account of experience phenomenologically adequate? Does it 'save the phenomena'? Do we have reason to believe that undergoing an experience of the species red is anything like the experiencing of a phenomenally red tomato? These may seem like silly questions. For the adverbialist already explained that the experience occurs in a red manner, i.e., that the experience belongs to a more determinate species of the genus *sensing,* namely, the determinate species *sensing redly*. Of course the adverbialist has said that. But may we not ask the adverbialist what it is that makes my experience be one of the particular species *red*. May we not ask what it is that makes my experience occur redly rather than greenly? In virtue of which of its intrinsic features does my sensing event belong to the determinate sensing species *red*?

I cannot help but think that a satisfactory answer to these questions must advert to phenomenal properties or qualia. My sensing event is of the species *red* or occurs in a red manner just in case it lets me experience or have a red quale.

This, it seems to me, must at least be part of the correct answer. Otherwise, I could presently be sensing in a red manner without experiencing phenomenal redness. And conversely, I might presently be experiencing phenomenal redness without the occurrence of sensing that belongs to the species *red*. But since the adverbialist theory of experience is supposed to offer an analysis of states such as experiencing phenomenal redness both of these possibilities must be excluded. If we proceed in this manner and let a sensing of the red species be one that allows the experiencer enjoy a red quale, then all my doubts about the phenomenal adequacy of adverbialism are dispelled. But the adverbialist would be well advised to resist this move. For this manner of understanding adverbially modified states of sensing brings back the problem the adverbialist set out to dissolve. The phenomenal properties are back and cry out for objects of some kind to support them. Since the adverbial analysis of experience was motivated by the desire to eliminate these objects, the adverbialist must reject the proposal to explain the notion of species of experience by appeal to the corresponding phenomenal properties.

What might the adverbialist say in a more positive vein? I cannot imagine what the adverbialist's positive account might look like. This may very well be due to the lack of imagination on my part. But the adverbialist need not give in, even if he cannot provide an answer to the question in virtue of which of its intrinsic features my sensing event belongs to the determinate sensing species *red*. For he might reject the question as not deserving of an answer. He might say that some sensings are of the red kind, others are of the green kind, yet others are of the blue kind, etc., and that is all there is to it. To even raise the question as to why or how this is or can be so is to misunderstand the basic or primitive nature of the facts involved. Thus, the adverbialist might deal with my second question not by answering it but by dismissing it as wrong-headed.

If that is what the adverbialist says, then the doubts about the phenomenal adequacy of the adverbial analysis of experience come back in full force. After all, the adverbial analysis of an experience of something red tries to get by without anything that is phenomenally red. Objects—be they external or internal, real or merely intentional—were dispensed with wholesale. So they are not what is red. Nor is the experience itself phenomenally red; nor, presumably, does it make sense to say that it is the experiencing subject that is red. In short, nothing about the event of sensing redly is phenomenally red. But surely the phenomenal fact with which the adverbialist starts is the fact that I experience the quality of phenomenal redness. That this fact be preserved (that this phenomenon be saved) is a condition of adequacy on any analysis of experience.

Given the adverbialist's refusal to further explain the nature of experiences that occur in a red manner, it remains an open question whether experiencing in

such a manner does preserve the phenomenal fact in question. And since the adverbialist's word of honor will not do to justify our belief that the adverbial analysis saves the phenomena, we should, I suggest, suspend our judgment about this matter. While this consideration falls short of a proof that adverbialism is false, it does direct us to investigate competing nonrelational analyses of experience.

10.4 Monism

Adverbialism is the only theory of consciousness that satisfies all of the following conditions: (i) it does not make the act/object assumption; (ii) it does not make the substance/attribute assumption; (iii) it currently enjoys some measure of acceptance. In the following sections I present a competing account of consciousness—the monistic theory of consciousness, or 'monism'—that satisfies (i) and (ii) while it fails to satisfy (iii) in a rather spectacular manner. I hope to remedy this shortcoming of monism by arguing that, unlike adverbialism, it satisfies an important fourth condition: it is true.

Monism is not a monolithic structure. Its various parts will be introduced as they are needed to answer the different questions that a theory of consciousness has to address. I begin by investigating how monism deals with the having-problem. This will force us to discuss the 'neutral theory of experience' (Section 10.5.1) and the bundle theory of the self (Section 10.5.2). The attempt to come to terms with the nature-problem requires the development of two additional parts of monism: the bundle theory of percepts (Section 10.6.1) and the 'other identity theory' (Section 10.6.3.1). These four main components of monism (together with various additions devleoped along the way) allow us to present monistic answers to the having-problem and the nature-problem.

10.5 Monism and the Having-Problem

10.5.1 The Neutral Theory of Experience

A theory of experience that does not make the act/object assumption is a nonrelational theory of experience. There is more than one way to overcome the act/object assumption. Adverbialism—the first nonrelational theory we encountered—dispensed with this assumption by, in effect, turning the intentional object (with its phenomenal properties) into a modification of the subject's act. What used to be the object is now the manner in which the act occurs. Adver-

bialism's merger of act and object proceeds at the expense of the object: it is absorbed into the subject's act, as it were. The resulting nonrelational theory of experience leans precariously towards the side of the subject. The neutral theory of experience (or 'neutralism') aspires to be a more balanced nonrelational theory of experience. It overcomes the act/object assumption by identifying the subject's act and the object of experience. By merging the subjective elements of the act with the objective elements of the object, neutralism ends up with a notion of experience that is as much subjective as it is objective. This is the sense in which the resulting notion of experience or sensation is neutral as between the act and the object it replaces.[140]

Why would anyone perform this rather unlikely identification of the act and the object of a sensation required by neutralism? First, our survey of the many failed attempts to come to terms with the having-problem suggests the hypothesis that these problems can only be solved by adopting a nonrelational theory of experience. I have argued that adverbialism, the standard version of the nonrelational theory, should be rejected. This forces us to take nonstandard versions of the nonrelational theory of experience seriously. Hence neutralism, with its initially implausible suggestion to identify act and object, now presents itself as the more promising alternative.

Second, these considerations can be backed up with a more direct argument in support of neutralism. The neutralist way of overcoming the act/object assumption has its roots in a skepticism about the subject. Since the act is the relation by means of which the subject becomes aware of the object, the act cannot survive the demise of the subject. Reasons to doubt the existence of the subject are therefore also reasons to give up the assumption of the act/object structure of experience. Skepticism about the self has been an enduring trait of the empiricist tradition. I will illustrate this rationale for giving up the act/object assumption by considering Russell's reasons for abandoning the self and for adopting neutralism.

Russell's reasons as follows. The subject must be given up because it cannot be observed. Hence, a relational theory of experience is untenable and the distinction between act and object cannot be upheld. Here is a passage that succinctly summarizes this thought:

> If there is a subject, it can have a relation to the patch of color, namely, the sort of relation which we might call awareness. In that case the sensation, as a mental event, will consist of awareness of the color, while the color itself will remain wholly physical, and may be called the sense-datum, to distinguish it from the sensation. The subject, however, appears to be a logical fiction…It is introduced, not because observation reveals it, but because it is linguistically convenient and apparently demanded by grammar…If we are to avoid a per-

> fectly gratuitous assumption, we must dispense with the subject as one of the actual ingredients of the world. But when we do this, the possibility of distinguishing the sensation from the sense-datum vanishes; at least I see no way of preserving the distinction. Accordingly the sensation that we have when we see a patch of color simply *is* the patch of color, an actual constituent of the physical world. (Russell 1978: 141-142)[141]

This consideration proves to Russell that "the patch of color and our sensation in seeing it are identical." (Russell 1978: 143) A question may arise as to how this talk of the *identity* of act and object is to be understood. Why not simply claim that what remains after the subject has been dispensed with is simply the object?[142] Compare: if we start out believing that Atlantis is west of Africa and subsequently discover that Atlantis is fictitious, then we do not proceed to claim that therefore Africa is identical with Atlantis. But it seems that this is how Russell wants to proceed in the case of the act and object of sensation.

I think that the talk of identifying act and object can be motivated in the following manner. The act may be fictitious; but the function that the act is supposed to perform is indispensable. The act is that which presents the sense-datum to us, that which brings us into contact with the sense-datum. Sense-data without corresponding acts would not be sense-data to anybody; thus, they would not be sense-*data* at all. This function of the chimerical act must be retained by whatever it is that remains after its the abolishment. For, if nothing plays this role of getting us into contact with the sense-datum, of revealing the sense-datum to us, then the sense-datum will not be a datum for us, it will not present itself to us. In the framework of the act/object model where the act of sensing is distinguished from the sensed object (i.e., the sense-datum) the sense-datum is 'presentationally inert.' It does not force itself upon us. We make it our own by reaching out to it with the act, i.e., by sensing it. If the existence of the act is denied, we no longer have a means to reach out towards the sense-datum, and we are doomed to live a life of inner darkness. To avoid this fate, the act's function must now be attributed to the successor concept of the sense-datum. Whatever it is that we get after abolishing the act/object distinction must, quite literally, present itself to us.

In order to avoid confusions that would otherwise be inevitable, we must find a new name for the successors of sense-data. Appropriating a term that Russell introduces in a slightly different context, we can call these successors 'percepts.' Adopting this usage, I can put the point as follows: upon discovering that the subject does not exist, Russell is led to identify the sensing with the sense-datum because he sees that the resulting percept has to play a role that combines the functions of the sensing and the sense-datum. The resulting notion of a percept is the notion of a self-presenting, qualia-bearing state or object. The percept's

ability to accommodate qualia will be discussed in Section 10.6.1. Its capacity to present itself is the topic of the next section.

10.5.2 The Bundle Theory of the Self

While the preceding considerations throw some light on the neutral theory of experience, they fail to show how this theory bears on the problems generated by phenomenal properties. What we have achieved so far is, at best, a rudimentary understanding of monism. Taken by itself, the neutral theory of experience does not help us to answer the having-problem or the nature-problem. In this section I introduce a further component of monism: the bundle theory of the self. When conjoined with the neutral theory of experience the bundle theory of the self allows us to formulate an answer to the having-problem. That is, we will be in a position to specify a relation that makes it intelligible why bearing it to a quale results in its being like something to do so. The phenomenal adequacy of this solution is discussed in Section 10.5.3. After raising some questions about the monistic answer to the having-problem (Sections 10.5.4 to 10.5.7), I turn to a discussion of the nature-problem in Section 10.6.

What is it to have a phenomenal state or a percept? One might be inclined to answer as follows: How can there be a problem about what it is to have a self-presenting state? Since a self-presenting state is a state that presents itself to the one whose state it is, your *having* a self-presenting state is just a matter of your being in such a state. And that is all that can and needs to be said about it.

I think that the reason why this answer does not satisfy is that the notion of self-presentation smells of the miraculous. I scent this odor too, and I do not know how to dispel it. Whether this is a good reason to reject the notion of a self-presenting state will depend on how well its competitor fares in this respect. What we should ask, therefore, is this: Is the notion of an act that presents a phenomenal object to us—and presents it to us in a way that results in there being something it is like to be us—any less miraculous than the notion of a self-presenting state? We are, of course, much more used to attribute such presentational powers to mental acts. But how the act is supposed to accomplish the remarkable feat of presenting us with a phenomenal particular in a way that lets us have the quale involved is a question that remains shrouded in darkness. This is the lesson that the arguments of Chapter Six and Chapter Eight teach. It is this notion of presentation that is troublesome no matter where we encounter it: in a *self*-presenting state or in an *other*-presenting act. Insofar as there is a miracle involved here, it embarrasses both the relational and the nonrelational accounts of experience. These considerations amount to no more than saying *tu quoque;*

but sometimes this is the best we can do. And in the case at hand it does seem to me that this kind of 'argument' succeeds in redistributing the burden of proof.

While the notion of a self-presenting state resembles that of an other-presenting act in involving the obscure notion of presentation, it differs in what we might call the logical structure of the presentation involved. If one motivates the neutral theory of experience in Russell's manner, viz., by denying the existence of the subject, then there is no self to which the self-presenting states (the percepts) can present themselves to. On the act/object theory of experience, on the other hand, the sense-datum is presented to the self via the act. Again it may seem that the neutral theory is worse off than the act/object theory of experience. For one might argue: presentation is always presentation *of* something *to* something; and a theory that requires there to be presentation of something to nothing is, therefore, not even false, but ill-formed.

This is a fair objection. But the monist can answer it, if not without considerable difficulty. The outline of the answer is as follows. The bundle theory of the self forms the centerpiece of the defense. Adopting this deflationary notion of the self, makes it possible to deny the absurd notion of the presentation of something to nothing without being committed to the sort of empirically undiscoverable self that philosophers in the empiricist tradition feel compelled to reject. Percepts present themselves to the self—a self that is no longer understood as a simple substance but as a bundle of percepts. For a percept to be presented to the self is for it to be a member of the bundle that is the self. Thus presentation is no longer conceived as something akin to the confrontation between two individuals. Given the bundle theoretic framework, presentation becomes a matter of membership in a unified group of percepts.

This, then, is what the monistic bundle theorist will say about the questions what it is for a percept to be present to a self and what it is for a percept to be self-presenting. For a percept to be present to a self is for it to be a member of an appropriate kind of bundle. And that the percept can achieve this presence to a subject via self-presentation is to say that it achieves this presence unaided by any activity of the other elements of the bundle of which it is a member. The mere membership of the percept in the bundle guarantees its being presented to the self that is the bundle. Thus, the question that opens this section can be answered as follows: to *have* the quale that is exemplified by a given percept is for that percept to be a part of the bundle you are.

10.5.3 The Phenomenal Adequacy of Monism

In giving the above reply to the having-problem I have carefully steered clear of the nature-problem. Since the nature-problem is prior to the having-problem, the

monistic account of how qualia are had must count as provisional. So far we have seen only two of the components of monism—neutralism and the bundle theory of the self. The final assessment of the monistic account of the having relation can only take place after monism has been sufficiently developed to address the nature problem. Having noted this proviso, I now want to elaborate my answer to the having-problem a little more fully by defending the phenomenological adequacy of the monistic solution of the having problem. In the previous chapters I have presented and criticized a variety of proposed solutions to the having-problem—among them the four representation-based accounts distinguished in Chapter Eight and the adverbialist theory of this chapter. Some of the criticisms were based on the particular features of these diverse theories and many of the points made turned on technicalities. But one criticism runs through all of these discussions like a thread. It is the complaint that these theories do not meet the condition of phenomenological adequacy. Phenomenological adequacy is the first requirement that any account of consciousness must satisfy. If 'trying on' the theory does not 'save the phenomena' the theory must be given up.

As far as I can tell, monism passes this test unscathed. It does not attempt to relocate the phenomenal properties onto objects that are either unfit to bear them or incapable of having the right kind of commerce with the experiencing subject; it does not reduce them away or eliminate them outright; it does not rely on the machinery of representation (of whatever order) to establish the very special relationship that exists between a person and the qualia she has; and it does not reconstruct the logic of phenomenal property terms in a way that raises doubts as to whether undergoing such logically revamped experiences would be like anything at all for the subject. Phenomenologically speaking, monism leaves everything as it is. The changes are in the ontology; the phenomenology stays the same. Monism tells a revisionary story about the structure of experience. But this restructuring of experience does not change what it is like to undergo experiences so conceived.

Perhaps it will be granted that monism avoids a panoply of mistakes and blind alleys that burden its rivals. But note that it is not alone in doing so. Every theory that says nothing or next to nothing avoids saying anything false. The 'we just do it' theories (of a broadly Moorean sort) are a case in point. Why, then, do I think that monism is more deserving of our attention? The combination of neutralism with the bundle theory of the self represents a radical and illuminating idea about how to capture the peculiar intimacy of the having-relation. It makes the qualia you have into a part of you. They (partly) compose you. They are (in part) what you are. It is not just a matter of pushing the qualia farther in, up close to the self, as it were. This way of putting it presupposes that you are one thing and your qualia are another thing and it so happens that these two things

are in close proximity. By dissolving the determinate boundary of the self, the present proposal overcomes the duality implicit in the previous suggestion. If you build selves out of qualia (among other things) then the question how the self gets to have its qualia no longer arises. On the contrary, it begins to seem inevitable that it would be like something to be a self that is made of qualia. Or so I say.

10.5.4 Challenging the Monistic Account of Having

I have argued that the combination of the neutral theory of experience with the bundle theory of the self provides a phenomenally adequate account of what it is to have qualia. I now want to address three problems that this proposal faces. First, one can challenge the monistic account of how qualia are had on the most basic level and claim that it is unilluminating. One can say that it remains unintelligible why having qualia in the manner outlined by monism results in its being like something to do so. Naturally, I do not share this impression. But I fear I cannot answer this objection in a way that will persuade the critic. One who feels that monism fails on this level will simply have to conclude that monism fails.

The second and third challenges attempt to exploit what are perceived to be two unwelcome consequences of the monistic notion of self-presentation. The second challenge can be stated as follows: self-presentation entails that we know our qualitative mental states infallibly. Infallibilism is false. Therefore, monism is false too. I reply by denying the first premise of the argument on the grounds that monism is a noncognitive theory of consciousness. Having a quale is not a cognition of any sort; much less is it a piece of infallible knowledge.

The third challenge runs as follows: self-representation commits the monist to silver-screen qualia. The naturalistically conceived world has no room for such properties. Hence we can embrace monism only if we are prepared to take leave from our cherished naturalistic sobriety. There is no short reply to this challenge. The monist's silver-screen qualia do indeed resist easy integration into the material world as it is commonly conceived. The discussion of the monistic solution to the nature-problem (Section 10.6) constitutes my long answer to the third challenge. I argue that you do not have to choose between monism and naturalism because an enlightened naturalism is permissive enough to accommodate the qualia of monism.

10.5.5 Challenge One: Monism Is Unilluminating

I suppose that a critic could try to do to me as I have done to all those discussed in the previous pages. He might say: "You simply assert that one has the qualia that compose one. But I do not see that. Is it not possible that a quale that

(partially) constitutes you is not, in any way, reflected in what it is like to be you? Your purported explanation of the having-relation is merely an explanandum in disguise." Though I see that someone might say these things I find myself unmoved—despite the fact that I cannot answer this objection. If you do not see that building the self out of qualia lets it have them, if this fact baffles you, then I cannot help you. By constituting the self out of qualia (among other things) the self and its qualia have been linked in the tightest possible way—a way, moreover, that naturally leads one to expect that a self and a quale so linked will result in there being something it is like to be the self having the quale in this manner. This is the point where my capacity to provide further explanations bottoms out. Every explanation one attempts, if pushed far enough, leads one to such ultimate premises—unexplained explainers, if you will. There is nothing intellectually disreputable in this. Nevertheless, it is frustrating to discover that one's interlocuter's sense of what constitutes an acceptable terminus of inquiry clashes violently with one's own. Here, for example, is one of Owen Flanagan's brute principles: "Some patterns of neural activity result in phenomenological experience; other patterns do not. The story bottoms out there." (Flanagan 1992: 58) The place at which Flanagan is content to bring his inquiry into consciousness to a close is none other than the place from which I want to start the same inquiry. On this point I find myself in complete agreement with Fodor:

> It is often argued, rightly in my view, that discovering the laws of brain/qualia correlation would not, in and of itself, solve the mind/body problem about consciousness. Barring an explanation of why the correlation is the way that it is, such laws would be surds; and only basic laws are allowed to be surds. (Fodor 1994: 122)

Such differences may prove to be irreconcilable. Upon reaching such an impasse, all that is left to say is that "one philosopher's…explanation is another's explanandum" (Cummins 1989: 65) and to leave it at that.

10.5.6 Challenge Two: Monism Entails Infallibilism

Percepts result from fusing the act (the sensing) with the object (the sense-datum). The object bears the quale. The act represents it. Thus the product of the fusion—the percept—combines two functions: it accommodates qualia (in a manner that still awaits clarification—see Section 10.6.1) and makes them available to the subject by being one of its elements. The percept bears the quale and reveals the quale to the subject of which it is a part. The percept does this unaided by presentations of any sort. This is the sense in which percepts are self-presenting. We might say that the percept needs no introduction—it announces itself and needs no help to make its presence felt. There is, then, in the percept

no distinction between its existing as an element in a bundle and its being experienced by the person who is this bundle; between one's tokening the quale-bearing percept and one's having the quale it bears. In this sense *esse est percipi* is true for percpets.[143] Embracing the controversial idea of self-presentation makes the monistic theory of consciousness powerful but also vulnerable. Those who reject the idea of self-presentation outright need hear no more about monism. But even those who are willing to countenance this notion must acknowledge that its employment appears to have undesirable consequences.

The second challenge to monism alleges that one such consequence of the self-presentation of percepts is infallibilism about one's qualitative mental states. The short route to the problem of infallibilism is as follows. If a quale's being instantiated in me and my being conscious of it are one and the same, then (i) I cannot fail to be conscious of a quale that is instantiated in me and (ii) I cannot be wrong when I am conscious of having a quale. So I am infallible about my qualia in the strongest sense of the term. I cannot fail to know about the qualia I instantiate; and I cannot have a wrong qualitative belief. Since the case against infallibilism is strong (see Section 5.7) so is the case against monism.

10.5.6.1 The Noncognitive Nature of Monism

The correct way of dealing with this problem is to simply deny that infallibilism is a consequence of the monistic theory. Monism is a noncognitive theory of consciousness. It holds that having a quale is not a matter of knowing or believing or perceiving or representing anything in any way. In short, to have a quale is not to cognize it. *A fortiori* it is not a matter of knowing anything infallibly. How, then, can this 'having' be characterized in a more positive vein? Perhaps we should stick with the minimalist answer that Schlick gave in the chapter he devotes to attacking the "so-called internal perception." There he states laconically:a "The [percept] simply *is* there, and this disposes of the whole matter." (Schlick 1985: 153) This is the least misleading way of talking about it. A large number of terms have been used in the attempt to render the pale having-idiom in a more palpable way: 'apprehend,' 'experience,' 'sense,' and 'feel' come to mind. All of these terms are misleading in suggesting an act/object structure of consciousness, a structure we have been at pains to avoid. Nevertheless, it is convenient to have a verb with which to describe the manner in which we enjoy our conscious mental states. I use the word 'feel' to play this role, always keeping in mind the nonintentional use to which it is put here. Of course this sense of the word 'feel' needs to be distinguished from the sense in which I feel for the doorknob in the dark, and from the sense in which I can be said to feel that causal theories of content are hopeless. What I would like to say, then, is that I feel a red quale when I look at a ripe tomato. Describing myself in this way I ex-

press what it is like to be me without implying that this feeling either consists in or entails a cognitive attitude towards the quale in question. This is the thesis of the noncognitive nature of consciousness.[144]

So far the noncognitive nature of monism has only been asserted but not supported. That consciousness is noncognitive is an important aspect of the present proposal. But it may be less than clear how this feature of consciousness drops out of the preceding account. The transition from the neutral account of sensation to the allegedly noncognitive nature of consciousness requires further comment. What underwrites the passage from 'nonrelational' to 'noncognitive' is a certain assumption about the nature of knowledge. Russell puts this point by saying that "there is a duality which is essential in any form of knowledge..." and that "generally, knowing is distinct from that which is known." (Russell 1975: 104) If one denies subject and fuses act and object, one is left with percepts that lack the structure—the duality—characteristic of a cognitive state. Thus the account of experience in terms of neutral (and hence nonrelational) percepts has the upshot that having a quale is a noncognitive event. We find Russell making this same move in the *Analysis of Mind:*

> It is of course undeniable that knowledge comes *through* the seeing, but I think it is a mistake to regard the mere seeing itself as knowledge. If we are so to regard it, we must distinguish the seeing from what is seen: we must say that when we see a patch of color of a certain shape, the patch of color is one thing and our seeing of it is another. This view, however, demands the admission of the subject, or act... (Russell 1978: 141)

The subject and the subject's act were of course already exorcised on the joint grounds of unperceivability and economy. This led to the identification of the sensation with the patch of color and from this it follows that the percept is noncognitive: "A patch of color is certainly not knowledge, and therefore we cannot say that pure sensation [i.e. a percept] is cognitive." (Russell 1978: 142)

Similar considerations lead Schlick to build his epistemology around the fundamental distinction between experience (*Erleben, Kennen*) and cognition (*Erkennen, Wissen*). Again it is the nonrelational character of experience that makes it noncognitive. Schlick argues that, if the knower were to become one with his object, he would not thereby attain some peculiarly intimate nonrelational knowledge. No, the knower who thus became one with his object would cease to be a knower: instead of knowing 'from the inside,' as it were, he would not know at all:

> There has been no dearth of philosophers who profess not to be happy with any lesser concept of knowledge than that of a complete union of the knower with the known...but the doctrine ought to have been rejected first of

> all on the ground that even if such a union were possible, it would not in any event constitute *knowledge*. (Schlick 1985: 81)

We can read this as a reaffirmation of Russell's claim that knowledge involves an essential duality. Thus, where this duality is missing knowledge is missing. This, Schlick holds, is the case in experience. Because we encounter this identity of subject and object in the case of experience, experience must be counted as noncognitive.

> Experience cannot be distinguished from experiencing and from what is experienced; it is all one and the same. For instance, a sensation of blue is an absolutely simple existent; one cannot separate within it the sensing of the blue and the blue that is sensed. (Schlick 1985: 137)

Because experience runs act and object into one, we must hold that "pure unelaborated perception or sensation is mere acquaintance (*Kennen*)." Schlick continues that as far as mere acquaintance or experience is concerned "it is entirely wrong to speak of a "perceptual knowledge." Sensation gives us no knowledge whatsoever of things, but only an acquaintance with them." (Schlick 1985: 87) Thus Russell and Schlick end up endorsing the noncognitive theory of consciousness for the same reasons: they hold that knowledge is essentially relational; they hold that what we have variously called 'experiencing,' 'sensing,' or 'feeling' is nonrelational; and they conclude that consciousness is noncognitive. And so should we. The nonrelational aspect of monism blocks cognitive accounts of qualitative consciousness. It is true that the self-presenting nature of percepts and their qualia—the *esse est percipi* of qualia—does guarantee something quite special: qualia are invariably felt by those who instantiate them. But to feel is not to know. Infallibilism is not even on the horizon.

10.5.7 Challenge Three: Monism Entails Nonnaturalism

According to the monist, percepts and their qualia present themselves to the subject to which they belong. If nothing else presents the quale to the subject, if the quale is itself present to the subject, then there is no room for a gap between the way the quale appears and the way the quale is. Thus the quale must be the way it appears. But—and this is the problem—self-presenting silver-screen qualia do not seem to fit into the world as conceived by sober naturalism. The Sellarsian grain problem that found its expression in the argument from atomism (Section 9.3) appears to block every attempt to integrate the simple and homogeneous silver-screen qualia into the material world. Hence, the qualia of monism can be retained only on pain of adopting a metaphysics designed under the influence. For the abstemious naturalist, this amounts to a *reductio*.

All the qualophiles we have discussed opt for celluloid qualia. That is, they opt for qualia that can be instantiated without being had. This allows them to construe qualia as well-behaved, nonthreatening properties. After all, the problematic 'what it's like' dimension enters the world only when an instantiated quale is being had. They are therefore free to make the quale itself into a quite ordinary property that a tomato (physical paint) or a neural state (mental paint) can readily exemplify. Of course, this leaves us with the rather large question how to have these very ordinary qualia in a way that makes the light of consciousness come on. But the friends of celluloid qualia face this challenge calmly. They have the answer—it is *representation*. Representing qualia lets you have them. There is much debate about the form this representation has to take. Is it perceptual or nonperceptual? Is it introspective or nonintrospective? Is it representation or metarepresentation? And so on. I have stated my reasons for rejecting this approach. But it may seem to have one great redeeming virtue: by turning qualia into nonmysterious properties, it makes it easy to find a place for qualia in the world.

In opting for self-presenting qualia I forego this elegant solution to the problem of finding a home for qualia. Self-presenting qualia are of the silver-screen variety—strange properties for which there is no place in the world of materialism, neither within nor without the head. So it may seem that the price of self-presentation is a 'spooky' metaphysics. I take this challenge very seriously. To rise to this challenge is to present a naturalistically respectable answer to the nature problem. In the following sections (especially in Section 10.6.3.1) I try to show that a certain naturalistically respectable framework—often thought to be a paradigm of philosophical tough-mindedness—is roomy enough to accommodate the problematic silver-screen qualia.

10.6 Monism and the Nature-Problem

The discussion of the questions raised by the monistic solution of the having-problem has naturally led to questions belonging to the nature-problem. In order to address the nature-problem we need to further extend monism. First I introduce the bundle theory of phenomenal objects (or percepts). This makes it possible to account for the exemplification of qualia in the absence of substantive qualia-bearers. The second crucial addition to monism is a lesser known version of the mind-body identity theory. It provides a metaphysical framework that combines naturalistic respectability with sufficient elbow room to accommodate qualia. When conjoined with considerations that place percepts and the qualia that constitute them into the head of the person who has them, the other identity

theory suggests the view that the brain is composed of qualia (and other things). This way of fitting qualia into the world tends to raise the worry of panpsychism. I therefore close the discussion of the monistic solution of the nature-problem with some remarks designed to deflect this concern.

10.6.1 The Bundle Theory of Percepts

The investigation of nonrelational theories of experience was motivated by the desire to avoid the difficulties encountered by the act/object analysis of experience. One of the main problems for the act/object model was the intractable nature of the qualia-bearing objects. None of the proposed construals of intentional objects provided satisfactory answers to the questions posed by the exemplification and the causation-problems. Some of the proposed objects were unfit to bear phenomenal properties. And those in which phenomenal properties could inhere could not enter into the appropriate causal relations with the experiencing subject. Thus, none of the candidate objects afforded a satisfactory home for qualia.

Adverbialism, the first of the nonrelational theories of experience, proposed a radical solution to this problem. Instead of offering yet another kind of intentional object for the phenomenal qualities to inhere in, the adverbialist rejects the way in which the problem is posed. The problem of phenomenal properties is not solved by discovering some special kind of qualia-bearing object. The all-important first step of the adverbialist solution of the qualia problem consists in the insight that it is a category mistake to take the predicates for phenomenal qualities to stand for properties of objects. By declaring these terms to stand for species of sensing, the adverbialist cleverly bypasses the difficult task of specifying a kind of object for phenomenal properties to inhere in. In place of qualia borne by objects, we now have acts of sensing variously modified. But I have argued that this ingenious strategy of dealing with the problems posed by qualia-bearing objects is flawed.

How does the neutral theory of experience fare on this account? In what manner do phenomenal properties enter into the percept? The percept results from fusing the sensing with the sense-datum (i.e., the object whose job it is to exemplify qualia). What happens to the qualia in the transformation of the sense-datum into a percept? So far, the ontology of percepts has not been addressed. We do not know to what ontological category percepts belong, nor do we know what their ontological structure is. If the monist answers that percepts are related to their qualia as substances to their properties, then he is confronted with the central problem that our discussion of intentional objects has revealed: the right kind of qualia-bearing objects do not exist. If the final word on monism is that

percepts are substances the attributes of which are phenomenal properties, then monism is in trouble.

But the monist can address this problem in a manner that parallels his solution of the having-problem. In keeping with his radical empiricism and his allegiance to the principle of ontological parsimony, the monist eschews substances wherever possible. The model upon which he construes percepts is not that of a substance with its attributes, but rather that of a bundle of properties. That is, rather than saying that a percept is a substance in which the phenomenal properties inhere, the monist will say that a percept just is the bundle of appropriately united phenomenal properties. Let us again turn to Russell in order to see how he motivates the bundle-theoretic account of phenomenal objects.

> Common sense regards a "thing" as having qualities, but not as defined by them; it is defined by spatio-temporal position. I wish to suggest that, wherever there is, for common sense, a "thing" having the quality C, we should say, instead, that C itself exists in that place, and that the "thing" is to be replaced by the collection of qualities existing in the place in question. Thus "C" becomes a name, not a predicate. The main reason in favor of this view is that it gets rid of an unknowable. We experience qualities, but not the subject in which they are supposed to inhere. The introduction of an unknowable can generally, perhaps always, be avoided by suitable technical devices, and clearly it should be avoided whenever possible. (Russell 1980: 98)

When, for example, one says 'this is red' while looking at the ripe tomato, one should not interpret one's statement as a subject-predicate proposition. For if one does that, "one finds that "this" becomes a substance, an unknowable something in which predicates inhere, but which, nevertheless, is not identical with the sum of its predicates." (Russell 1980: 97) The correct analysis of the statement yields the following results:

> That "this is red" is not a subject-predicate proposition, but is of the form "redness is here"; that "red" is a name, not a predicate; and that what would commonly be called a "thing" is nothing but a bundle of coexisting qualities such as redness, hardness, etc. (Russell 1980: 97)

It is not difficult to see how adopting the bundle theory helps solve the exemplification-problem. On the bundle theoretic account, the question how a given substance can instantiate a given property simply does not arise. In this respect the bundle theory rivals the adverbialist solution in elegance. Both accounts make this problem go away, albeit in quite different ways. According to the monist, the home of the quale is the percept and the quale inheres in the percept as an element does in a bundle and not as a property does in a substance.

What we have here is yet another theory about what it is for a phenomenal property to inhere in what it qualifies. We have already looked at two other theories of what it is for something to exemplify a quale. The bundle theory is the third and last such theory that I want to consider. We started out with an act/object model of experience and were led to the question whether intentional objects are suitable bearers of phenomenal properties. This discussion proceeded on the assumption that phenomenal properties are to be taken as attributes of underlying substances. The difficulties encountered with the substance/attribute assumption led us to consider the adverbialist theory of consciousness. On this theory, phenomenal properties are construed as species of sensing events. The latest proposal, the bundle theory, explains the inherence of a phenomenal property in what it qualifies as being a matter of the joint manifestation of this and all the other phenomenal properties that, as we would ordinarily say, the object has. The substance that is supposedly needed to support all these properties is eliminated as superfluous.

The bundle theory is a happy combination of what is best in its two rivals: it preserves the good features of the substance/attribute theory and of the adverbial theory, while inheriting the problems of neither. What is good about the substance/attribute theory is that it treats phenomenal properties as just that: properties of things. On that account the adverbial theory fails. The role it assigns to phenomenal properties remains obscure. What is good about the adverbial theory is that it obviates the need for qualia-bearers that exemplify the phenomenal properties. On that account the substance/attribute theory fails. Its need for substrata, for supports of phenomenal properties, is a serious liability. The bundle theory is a most commendable mixture of the two. It is like the substance/attribute theory in treating phenomenal properties as genuine properties. And it resembles the adverbial theory in not requiring offensive substrata. The bundle theory delivers genuine phenomenal properties freed of the burden of murky substances.[145]

Theories that make notoriously difficult metaphysical issues seem easy rightly arouse our suspicion. Their great power is often discovered to be a symptom of a concealed incoherence. The bundle theory may be a case in point. My hope is that the metaphysicians will prove it innocent. But one reason for doubting its coherence I want to address here. The main problem with the bundle theory may seem so obvious that no metaphysician is needed to discover it and so serious that no metaphysician can fix it. The challenge can be put as follows: the idea of property instantiations without property instantiators makes no sense. The supporting substances at the core of property clusters are not gratuitous inventions of malicious metaphysicians; they are required by the very notion of property exemplification. We cannot understand the notion of an 'unsubstantiated' prop-

erty exemplification. How should one reply to this challenge? I fear the best I can do is to admit the obscurity and to point out a parallel obscurity in my opponent's position. Admittedly, a property instantiation unsupported by a substance is a queer thing. But so is a property instantiation that is supported by a substance. The support-metaphor can only bear so much weight. It must not lull us into believing that we *understand* how something can exemplify a property by 'standing under it.' Properties are not like table cloths; nothing about them suggests that they collapse pitifully absent a sturdy supporting structure. These remarks do not answer our opponent. But they may give him pause by pointing out the generality of the problem he has raised. The nature of property instantiation is one that everybody has to deal with. On this count, at least, the neutral theory of experience is no worse off than the rest.[146]

How, then, does the monist propose to house the qualia? The phenomenal red that I experience when visually experiencing a tomato inheres in the object that is the bundle of phenomenal properties as which I experience the tomato. That is, the phenomenal red is a member of the bundle of compresent phenomenal qualities that together constitute the tomato as I experience it. Thus, there is a sense in which the monist can acknowledge that the phenomenal red of the tomato is borne by an object. However, the object is not a problematic substance or substrate, but an innocuous bundle of phenomenal qualities. It is a substance only in the metaphysically harmless sense of being an object of which we can say that it has (phenomenal) properties. This is how the proponent of the neutral theory of experience proposes to house qualia without getting stuck with problematic qualia-bearers.

10.6.2 The Physical Location of Percepts

We have learned that and how qualia are in the percept. But this still does not tell us the full story about the location of qualia as long as the location of percepts is unknown. The above quotations from Russell do not speak to the question of the location of the percept in the larger scheme of things. In the present context talk of the location of percepts is interestingly ambiguous.[147] A question about the location of percepts may be a question concerning their physical (or spatial) location—a question that can be answered by the specification of a few coordinates. But a question about the location of percepts may also be a question concerning their metaphysical location—a question about where the percepts fit into one's overall ontology. The answers one gives to these questions will be connected. It may, nevertheless, be helpful to begin by attempting to answer these questions severally. At this point I will try to say something about the spatial lo-

cation of percepts. In Sections 10.6.3-10.6.3.3 I try to find a place for percepts in a broader metaphysical framework.

A number of considerations persuade the monist to locate percepts in the percipient's brain. This result seems to be a direct consequence of neutralism's identification of act and object. The considerations that carried the most weight with Russell are grounded in a scientifically based account of perception. If one acknowledges—as Russell thinks we must—that perception is a causal process, we have no choice but to locate 'both' the seeing and what is seen in the brain of the percipient.

The first consideration is grounded in the monist's neutral account of experience. In the neutral theory act and object, the sensing and the sensed, are fused into one to yield the percept. This identification places significant constraints on the possible locations of the percept. Among those committed to the act/object model there has been considerable dispute about the proper location of the object. The location of the act has not given rise to corresponding disagreements. The act has, for the most part, been located internally—in the mind/brain or in the Mind. The monists agree that this is where the act belongs. But for them the consequences of this choice are momentous. It means that the object too is internal, for they identify act and object. Thus the percept, in its entirety, is located internally.

The second set of considerations is grounded in scientific considerations. I will present two variants of this idea. The first approach is grounded in a certain understanding of the nature of science. It is vividly expressed by Max Planck, whose thought exercised considerable influence on paradigmatic monists like, for example, Schlick and Feigl. "The first requirement of scientific thought," Planck writes, "is to acknowledge and to carry out the separation between the outer and the inner world." (Planck 1990a: 99)[148] Following this injunction in one's scientific work leads to the "emancipation from the anthropomorphic elements, especially the specific sense impressions." (Planck 1958a: 9) Thus one ends up with a unified scientific world picture of ever increasing empirical accuracy and ever decreasing graphicalness (*Anschaulichkeit*). "Indeed, when compared with the original naive world picture, the present scientific world picture presents a peculiar sight that seems downright outlandish. The immediately experienced sense impressions…have completely vanished from the world picture. Seeing, hearing, and touching are not mentioned in it." (Planck 1990b: 243) In thus withdrawing from the world of the senses, "the scientific world picture retreats ever farther into the abstract; purely formal operations play increasingly important roles and differences in qualities are progressively reduced to quantitative differences." (Planck 1958b: 184) In this austere world picture percepts are marginalized. The best they can hope for is a place in the 'dustbin of the mind.'

This is exactly the thought we find expressed in the monistic literature. Grover Maxwell, for example, presents the following approving summary of Russell's discussion of this idea:

> If our current theories in physics, neurophysiology, and psychology are at all close to the truth or even if they are at all headed in the right direction, then a complete description, including a complete causal account, of everything that is involved in perception except the private experience itself would mention only such entities and events as submicroscopic particles, electromagnetic quanta, etc., and their relations and interactions with one another and with, for example, neural termini in the retina, afferent neural impulses, and patterns of neuronal activity in the brain. At no point in the entire, complete description and causal explanation is there mention of any first order property such as colors *until* we come to the private experience that results from the pattern of neuron firings in the brain. It seems to me that we must conclude that colors are exemplified *only* in our private experiences and that there is no reason to believe that they are ever properties of the material objects of the external environment. *What holds for colors must also be true of all of the first order properties that we perceive directly.* (Maxwell 1970b: 19)

While the preceding consideration is a staple of empiricist thought on this matter, Russell is fond of presenting a variant of this idea. The crucial consideration that seems to persuade him to locate percepts internally—'in' the brain—is the causal theory of perception. What places the percept into the brain is its position in the causal chain of events that constitutes the perception of a given physical stimulus:

> Whoever accepts the causal theory of perception is compelled to conclude that percepts are in our heads, for they come at the end of a causal chain of physical events leading, spatially, from the object to the brain of the percipient. We cannot suppose that, at the end of this process, the last effect suddenly jumps back to the starting-point, like a stretched rope when it snaps. (Russell 1954: 320)

Note that his argument does not turn on assuming the 'colorlessness' of the world of physics. This strengthens the argument allowing it to work even if color is a feature of the extramental world. And this is a genuine possibility. For, as has often been noted, the argument that Maxwell summarized above does not entail that color is not a part of the world of physics.[149] Of course the argument from the causal theory of perception is not free of assumptions. The two most salient ones are these. First, it assumes that causal order determines spatial order. This principle places the event of seeing spatially into the brain at the end of the causal sequence that leads from the external object through the afferent nerves into the visual center of the brain. Second, it assumes the neutral theory

of experience. Were it not for this second assumption, it might plausibly be argued that the causal theory of perception places the act (or the vehicle of perception), *but not its object,* into the brain. I think that this criticism—a criticism leveled against Russell by nearly every philosopher who even bothered to discuss this argument—overlooks the crucial fact that the monist operates with a nonrelational theory of experience. If the sensing is in the brain then so is the sensed, for they are one and the same. Hence the admission that the causal theory of perception places the act into the brain amounts to the admission that it places the percept into the brain.

Placing the percepts in the brain has the consequence that "the whole of our perceptual world is, for physics, in our heads." (Russell 1954: 336) Russell distinguishes between "the physical world of physics, and the physical world of everyday experience." (Russell 1956a: 141) The latter is composed of everything you see, hear, touch, smell, etc. Thus the world, as it presents itself to me in experience, "is part of my mental life" (Russell 1956a: 141) and exists only in my head. So far as the constructed unintuitive space of physics is concerned, the three-dimensional space of your experience is at the 'point' at which your head is located.[150] The full force of this view becomes palpable when applied to quite mundane examples. Russell does not shy away from spelling out these consequences: "If physics and physiology are to be believed, I do not "see" the furniture in my room except in a Pickwickian sense." (Russell 1976: 221) There is only one thing that you see in a non-Pickwickian sense and that is your brain. "I maintain" Russell tells us "that we *can* witness or observe what goes on in our heads, and that we cannot witness or observe anything else at all." (Russell 1975: 19) All of this makes it quite clear where Russell—our representative of monism—wants to put the percepts. Percepts are in the brain. Qualia are in percepts. So the monist locates qualia in the brain. This seems to answer the location question in its first interpretation, i.e., interpreted as a question concerning the spatial location of qualia.

But unfortunately we cannot stop the investigation here and declare that the monist has provided qualia with a home. Two considerations (that we have already encountered earlier) seem to show that this answer to the location question is plainly wrong.

First, if one moves percepts into the brain one encounters the colored-brain problem (see Sections 2.2.2.8 and 7.5.2). One need not take it on Dretske's authority that qualia "are nowhere to be found in the head." (Dretske 1995: 151) In-depth studies of the brain confirm what a casual glance at a living brain suggests: the brain does not exemplify the qualia of the experience for which it is responsible. So, the monistic answer to the location question is wrong. Monism is a richer theory than HRT—we now have percepts, self-presentation, and the

bundle theory of the self and the phenomenal object. But these additions do not seem to change the fundamental fact that the brain does not exemplify the qualia one experiences. If monism is to be saved, the monist must explain how the silver-screen-qualia of monism can be in the brain in a way that conceals them from the sight of the observing neurophysiologist. The next section is devoted to a discussion of this problem.

Second, if one moves percepts and their qualia into the brain one has to face up to the problem raised by the third challenge to the monistic account of the having-problem (see Section 10.5.7). Sellarsian worries about grain and the argument from atomism to which they give rise (see Section 9.3) seem to show that silver-screen qualia do not agree with matter. One who insists on working with silver-screen qualia must face the argument form atomism head-on. Unlike the friend of celluloid qualia he cannot strip his qualia of the problematical features that block their identification with naturalistically well-behaved properties of physical objects. If one holds that the quale is the way it seems then one cannot hope to make one's qualia more manageable by distinguishing the way the quale is from the way it appears. If monism is to be saved, the monist must explain how the simple and homogeneous silver-screen qualia can inhere in the grainy and inhomogeneous matter of the brain. I regard this as the most serious challenge to monism. In Section 10.6.3 I try to show how the monist can meet it.

10.6.2.1 The Colored-Brain Problem

Dretske and many others attach enormous weight to the "obvious fact" (Dretske 1995: 151) that qualia cannot be observed in the head of the one who enjoys them. This consideration, more than anything else, is responsible for the relocation$_w$ of qualia and all that this entails. The colored-brain problem can be stated even more forcefully. It is not just a matter of not finding the red in the brain of one who looks at a ripe tomato. The problem is that we do seem to find qualitative features in the brain of the tomato's admirer and these features exclude the red quale from being instantiated there. Simply put, the brain of one who looks at a ripe tomato is grey. What is grey is not red. Thus the red that the tomato's observer sees cannot be instantiated in his brain.

The monists are not impressed by either version of the argument. Their writings contain numerous discussions of neurophysiologists who peer into living human brains and of what they do or do not see there.[151] The monists take this example as teaching a number of lessons. One of them is that the colored-brain problem is not a problem for monism. On the contrary, reflections on the observations made by neurophysiologists strengthen the monistic position.

The thesis that percepts and their qualia are located in the brain is what gets us into this problem. But it is also what gets us out of this problem. For it means that what the neurophysiologist sees is in his own brain, not in the brain he is examining. The brain he is examining may be hosting a red quale. But this red quale the neurophysiologist does not get to have. The only qualia that he can feel are the ones that his own brain instantiates. And there is no reason to think that the observation of a brain that instantiates a red quale will lead to the instantiation of a red quale in the observer's brain. This may not be immediately obvious in the case of color qualia. But consider auditory or olfactory qualia, for example. Surely, we do not expect to hear the plaintive tunes of "Georgia on my Mind" when we put our ear up close to the brain of one who is hearing this song. If your psychiatrist were to propose to 'listen to your thoughts' by using a sensitive stethoscope applied to your brain, you would switch doctors. Thus, the fact that the brain of one who sees red does not look red to its observer is just what the monistic theory predicts. And this fact has no tendency to show that the red quale is not in the grey looking brain you are observing.

Those impressed by the colored-brain problem assume that the red and the grey quale must be allocated to one brain—the brain of the patient. This creates a problematic buildup of qualia in the patient's brain. The relocationist's$_w$ solution to this problem consists in the radical proposal to remove all qualia from all brains. But we found a more conservative solution. By carefully redistributing the qualia over the brains of the patient (red) *and* the neurophysiologist (grey) we have sufficiently enlarged available brain space to keep all qualia inside.

This will serve as an explanation of why even the most intense scrutiny of a brain does not turn up the qualia it harbors. In asking this question one betrays a fundamental misunderstanding of the perceptual process involved in the neurophysiologist's search for qualia. Such a search can only lead the searcher to find—i.e., to have or feel—her own qualia. The assumption that the scrutiny of a brain would lead to an observation of the qualia it contains is groundless. Hence, the neurophysiologist's failure to discover the relevant qualia in the patient's brain proves nothing. And the apparent discovery of an incompatible quale (viz., the grey one) in the subject's brain rests on a confusion. The real location of that quale is in the observing and not in the observed brain.

We have seen how much weight Dretske attaches to the colored-brain problem. The 'obvious fact' that the qualia someone has cannot be observed in her head is the mainspring of his relocationism$_w$. Even so, he seems to have felt a flicker of doubt as is borne out by the following wonderfully suggestive remark: "If there are sensations and feelings in there [in the head], they seem to be well camouflaged." (Dretske 1995: 36) Our discussion of the colored-brain problem has shown that the camouflage metaphor is a little misleading. The reason we

cannot observe the qualia of others is not their skillful camouflage but the fact that we are doomed to keep looking in the wrong place. When you try to look at my head you end up looking at our own. There is, then, a sense in which Dretske is right. Strictly speaking, nobody can observe qualia (or anything else) in another's head (or in anything else that isn't her own head). This is a consequence of the internalization of percepts. The resulting view may strike Dretske and many others as absurd. But this is a quite general issue concerning the theory of perception, not the much more specific difficulty raised by the problem of the colored brain. And this more limited problem the monist can address.

10.6.3 The Metaphysical Location of Percepts

The monistic answer to the colored-brain problem is fine as far as it goes. But it does not go very far. For it makes the problematic assumption that brains can exemplify silver-screen qualia. In this way the discussion of the colored-brain problem has led us right back to the (by now familiar) grain-based objection to monism (see Section 10.5.7 and page 292). Unlike the colored-brain problem, the problem articulated in the argument from atomism is a genuine difficulty. Nothing short of a radical reassessment of the place of percepts and their qualia in the metaphysical order of things will allow us to address this problem. So I turn to a discussion of the metaphysical location question.

10.6.3.1 The Other Identity Theory

We need to find a metaphysical framework that allows us to see how the simple, homogeneous, silver-screen qualia of monism fit into the material world. The task is made more difficult by the requirement that this framework has to satisfy two conditions that appear to be in considerable tension. First, the framework must be roomy enough to accommodate the peculiarities of the monistic account of consciousness. Second, it must live up to the standard of naturalistic sobriety. I think that the identity theory provides the metaphysical framework we are looking for. This may come as a surprise. For the identity theory, as commonly understood, is a paradigm of philosophical tough-mindedness. That is, it is the distillation of materialist and reductionist trends in the philosophy of mind. It is exactly the sort of framework in which there seems to be no room for the percepts and silver-screen qualia of monism. This impression is not without foundation. The version of the identity theory that has come to dominate current thought is very close to the stereotype just sketched. But I will argue that there is a second version of the identity theory that fits our bill. I will call this latter version of the theory 'the other identity theory.'[152]

Any version of the mind-body identity theory faces the problem that the mind-body identification is counter-intuitive, especially so in the case of qualitative states. Any version of the identity theory will, therefore, have to tell a story that helps us overcome this enormous intuitive obstacle. The two versions of the identity theory employ diametrically opposed strategies when dealing with the intuitive implausibility of the identification. Reflection on this contrast helps to bring into clearer focus the difference between the mainline identity theory and the other identity theory. Both parties agree that something has to give, if the identification is to go through. But there the agreement ends. The mainliners hold that the mind must give. The proponents of the other identity theory hold that matter must give.

Let us begin by reminding ourselves of the relevant features of the mainline version of the identity theory. Our primary interest is, of course, how the identity theorist deals with qualia. This question we have already answered. For Armstrong is the leading light of this school of thought and we have discussed his views at length. Armstrong works with celluloid qualia that are wholeheartedly relocated onto the external, physical object of perception.[153] There he reduces them to, and identifies them with, complexes of primary properties on the surfaces of these objects. These qualia are had by being made the object of an (external) perception that itself becomes the object of an internal perception. Note that the special emphasis on qualia lets us see the identity theory in a somewhat unusual light. Ordinarily the focus is on the identification of mental states with neurophysiological states. But our interest in qualia together with the relocation$_w$ move shifts our attention away from the neurophysiological states in the head towards the surface properties of external physical objects. For the properties with which Armstrong identifies qualia are out there on the object and not on the neurophysiological state that represents the object. Only when we turn our attention to the having-problem do mental states begin to play a role.

Armstrong's version of the mainline identity theory identifies qualia with surface properties of external physical objects and the mental that represent and metarepresent these qualia with brain states. This means that Armstrong has to deal with *two* counter-intuitive identifications. Let us first look at the identification of mental states with brain states. As we have seen (see Section 9.3.2), Armstrong turns mental states in general and qualitative states in particular into mere 'that whiches' that are 'all gap.' He does this by singling out the alleged qualia of mental states as the problematic properties that block the identification. He then peels the qualia off the mental states, moves them onto the external object, and identifies them with certain of its surface properties. Freed from these ill-behaved properties, mental states turn into pliable 'that whiches' offering no resistance to the identification with neurological states. Once this purely struc-

tural view of the mind is embraced, its identification with the brain no longer poses an intuitive problem.

Note that this way of dealing with the problematical mind/body identification solves one problem by creating another one of equal difficulty. For now the qualia are out there, on the object, where they seems quite out of place. How, then, does Armstrong propose to sweeten the identification of qualia with primary quality complexes on the surfaces of physical objects? His treatment of the relocated qualia parallels the one accorded the mental states that represent them. Recall, for example, what Armstrong has to say about the quality of redness (see Section 9.3.2). By suggesting that "our concept of red is *all* blank or gap," (Armstrong 1968: 275) he primes redness for a smooth reduction to the primary quality complex with which he wants to identify it. The impression that redness does not confront us as 'all gap' is due to the illusion of concrete secondary quality. The features of qualitative experience that do not naturally drop out of this model—homogeneity and simplicity, for example—are thus declared as illusory additions.

Armstrong employs the same strategy in both cases. The relocated$_w$ qualia as well as their mental representations are 'thinned out' into mere 'that whiches' that do not resists identification with the naturalistically respectable properties of external physical objects and brains states respectively. This is the sense in which the mainline identity theory insists that the mind and its relocated$_w$ qualia have to give in order for the mind/body identification to succeed.

The mainline version of the identity theory does not recommend itself to the monist. Its account of what qualia are and of how they are had contradicts monism in every single point. If the other identity theory is to afford a metaphysical framework for the monistic theory of consciousness, it better be 'very other.' I hope that even a brief exposition of the alternative view will show that it stands the mainstream version of the identity theory on its head. The mainline theorists hold that the mind has to give. The other identity theory pursues the opposite course: matter has to give. The theory emphasizes the fact that matter is known to us only as a mere 'that which' of whose intrinsic nature we are ignorant. This claim is in marked tension with the apparent fact that perception acquaints us with all sorts of intrinsic properties of matter. But after it is realized that these alleged properties of matter belong not to the material objects we seem to see but to our seeing them, matter turns into something known only in a quite abstract and structural way. Once this 'emptiness' of our notion of matter is grasped, the identification of the qualitative with the material no longer poses an intuitive problem. For all we know, the inner nature of the structure known to us as the material brain may be qualitative.

It is now time to elaborate some of the features of the other identity theory. I will, for the most part, follow Herbert Feigl's lead. Ironically, Feigl's prominence as an early identity theorist can in part be traced to the fact that his view has often been uncritically lumped together with the much better known mainline version of the identity theory. Thus, some of Feigl's fame rests on his being associated with a view that he never held. This is especially regrettable because it tends to obscure the existence of, and Feigl's important contributions to, the other identity theory. The other source upon which I will rely is Russell. His guiding influence was palpable throughout the development of the monistic position. I believe that Russell does supply a metaphysical framework within which the monistic theory can thrive. But Russell's later work in this area has tended to go unnoticed or, when noticed, to cause considerable confusion. To a large part this can be attributed to the difficulties involved in understanding Russell's doctrine of neutral monism. Since my concern is the development of a metaphysical framework for the monistic theory of consciousness and not the interpretation of Russell's later philosophy, I will steer clear of the interpretive disputes surrounding Russell's metaphysics of neutral monism. Instead, I will use such bits of Russell's doctrine as are useful in developing my own project.[154]

10.6.3.2 Matter as All Gap: Structural Realism

The monist locates percepts in the head and takes that to entail that the "physical world of my everyday experience is part of my mental life." (Russell 1956a: 141) Keeping this doctrine in mind is the key to an understanding of the various aspects of monism. We have already seen how this doctrine allows the monist to deal with the colored-brain problem which arose in the context of the discussion of the physical location of percepts. We will now see that this doctrine also casts light on the monist's attempt to locate percepts metaphysically.

In moving the percepts inwards the monist effects what one might call an epistemic revolution. You may think that it is easy to acquaint yourself with, say, a tomato. One look is all it takes. And you may also think that it is hard or even impossible to acquaint yourself with internal tomato representations. No matter how hard you try to look inwards you end up looking at the external tomato again. According to the monist, this gets things backwards. Acquainting yourself with the tomato percept is easy. It happens every time you 'take a look at the tomato.' But acquainting yourself with the tomato is impossible. As Russell put it succinctly: "I know about what is happening in the brain exactly what naïve realism thinks it knows about what is happening in the outside world." (Russell 1979: 104)

The knowledge that monism takes away with one hand it gives back with the other. Both of these moments play an important role. Let us first look at the consequences of decreasing one's cognitive hold on the external world. By severing the links of acquaintance with the objects in the external world, monism turns them into unobservables. We infer their existence as the causes of our perceptions and thus get to know about them by description only. Such features as they may have in addition to the ones required by their causal role remain beyond our grasp.[155] Thus we end up with a very sketchy conception of the objects in the external physical world. We know them only as somethings of a sufficient structural complexity to produce in us the percepts we enjoy. What they are in themselves we do not know. But the monist welcomes this ignorance. The gappier the physical becomes the more likely it is to have room enough to accommodate the qualia.[156]

Notice that the fate that has befallen the external physical objects in the monistic regime is quite similar to what the mental states had to suffer at Armstrong's hand. They turned into ghosts of sorts, determinate only with respect to their structural/causal features but completely diaphanous as regards the rest. The point of so emaciating a given set of objects is, of course, also the same in both cases. Something reduced to such a tenuous condition—a mere 'that which'—puts up no resistance when subjected to identification with objects of a more robust constitution.

We have now before us a rough picture of what Maxwell has called structural realism. It is a form of realism because it acknowledges the existence of a mind-independent reality. It is structural because it holds that our knowledge of this reality is confined to its causal skeleton. We do not know the intrinsic nature of this reality—we do not know what it is that exemplifies the known structures. Our ignorance is profound. Not only do we not know what the intrinsic nature of these structures is—*we do not even know what it is not.* This opens up possibilities foreclosed by more confident epistemologies. In this way, the ignorance resulting from the retreat of the percepts into the head gives us an unprecedented freedom to speculate about the nature of the physical.

10.6.3.3 Filling the Gap: Qualia as the Basic Reality

By weakening our hold on the physical objects in our environment, the monistic internalization of percepts undermines the belief that we have direct perceptual knowledge of the material world. But by bringing us into the closest possible contact with our percepts, the monistic internalization of percepts suggests a provocative hypothesis about the nature of matter. In the previous section I discussed the first consequence of the internalization move. Now I turn to a discussion of the second.

I have already spelled out how I propose to understand the idea of the close contact between us and our qualia (see Section 10.5.2). Although Feigl is not committed to the details of that account, the manner in which he tries to exploit the idea of closeness is quite compatible with what I have said there. Instead of self-presentation, Feigl talks about the coincidence of evidence and reference in the domain of the qualia we feel. In the case of qualia there is no appearance/reality distinction. In feeling a quale we feel the quale itself, not merely some representation of it. This is the only point where we come into direct contact with reality as it is in itself. Where we do not have this coincidence of evidence and reference—the coincidence of how things appear and what they are—we do not know the nature of the evidenced reality. The failure to respect this crucial distinction blinds one to the possibility of the other identity theory:

> Even sophisticated analytic philosophers tend to confuse the meaning of physical concepts with the perceived or imaged appearance of physical things. No wonder then that we are told that the identity of certain neurophysiological states (or features thereof) with raw feels is a logical blunder. If the denotatum of "brain process (of a specified sort)" is thus confused with the appearance of the grey mass of the brain as one perceives it when looking into an opened skull, then it is indeed logically impossible to identify this appearance with the raw feels, e.g., of greenness or of anxiety. (Feigl 1958: 454)

If one resists the inclination to attribute the features of the evidence (the appearances) to the evidenced (physical states, neurophysiological states, etc.), their abstractness stands revealed. The theoretical conceptual system of the sciences—here Feigl includes physics, biology, psychology and the social sciences—"is in principle non-intuitive (*unanschaulich* as the Germans call it, i.e., unvisualizable)." (Feigl 1958: 454) Cleansed of the pictorial root flavor (see Feigl (1958: 455-456)) imparted by their evidential basis, scientific concepts acquire a degree of transparency that makes the identification of the physical with the qualitative seem possible.

It should now be obvious why Feigl claims that "the central core of the proposed solution rests upon the distinction between evidence and reference." (Feigl 1958: 466) By carefully attending to the roles that evidence and reference play in different domains, Feigl establishes two crucial claims of the other identity theory. First, the recognition of the coincidence of evidence and reference in the restricted domain of qualitative experience is what opens up this special window onto reality as it is in itself. Second, the meticulous distinction of evidence and reference in all other domains reveals the abstractness of our scientific concepts. This is how Feigl puts these two ideas together to arrive at the intended identification of the qualitative with the physical:

> The physical sciences consist of knowledge-claims-*by-description*. That is to say that the objects (targets, referents) of such knowledge claims are "triangulated" on the basis of various areas of observational (sensory) evidence. What these objects are acquaintancewise is left completely open as long as we remain within the frame of *physical* concept formation and theory construction. But, since in point of empirical fact, *I* am directly acquainted with the qualia of my own immediate experience, I happen to know (by acquaintance) what the neurophysiologist refers to when he talks about certain configurational aspects of my cerebral processes. (Feigl 1958: 450)

Feigl's proposal has often been misunderstood in a way that hides its radicalism. Feigl is *not* suggesting that the reality designated by the neurophysiological terms may also *appear* to the experiencing subject in the guise of qualitative experience. This way of putting it demotes the qualitative into a mere appearance of an underlying reality—a reality that may well have the reassuring solidity of matter as traditionally conceived. On this misreading, Feigl's view can be interpreted as a variant of the mainline version of the identity theory. Perhaps the inclination to read Feigl in this way is grounded in a charity principle demanding the least wild interpretation of a philosophical doctrine. Making Feigl into a mainline identity theorist may be charitable, but it is false nevertheless. Taking seriously the thesis of the coincidence of evidence and reference—of appearance and reality—in the domain of the qualitative excludes this proposed interpretation. In this special case the evidence we have is not evidence for something else whose inner nature remains hidden. In the realm of qualitative experience there is no reality behind the appearance because the appearance is the reality. Feigl explicitly repudiates the above interpretation and insists that it misses the unique nature of the mind-body identification: "We don't have two kinds of evidence for one and the same entity (event, process, etc.). In direct acquaintance we *have*, we *experience* the datum (it is not evidenced, it is *evident!*), and we identify it with a physical process." (Feigl 1958: 452-453) Feeling qualia acquaints you with the reality that neurophysiology describes. As Feigl put it: "The data of experience *are* the reality which a very narrow class of neurophysiological concepts denotes. (Feigl 1958: 453) And again: "The directly experienced qualia…are the realities-in-themselves that are denoted by the neurophysiological descriptions." (Feigl 1958: 457) Everyone is familiar with the slogan of the mainline version of the identity theory: the mind is the brain. The slogan for the other identity theory might read: the brain is made of qualia.

Jerry Fodor has written: "If aboutness is real, it must really be something else." (Fodor 1987: 97) Many self-professed qualophiles seem to be committed to an analogous maxim about qualia: If qualia are real, they must really be something else. Hence the great eagerness to substitute some form of celluloid qualia for the old silver-screen qualia. The various proposals put forth in the

spirit of this maxim share two features. They succeed in taking the mystery out of qualia by making them into quite plain properties—so plain, in fact, that it remains mysterious how these humble ingredients are supposed to yield a phenomenologically adequate account of qualitative consciousness. The other identity theory is conceived in a quite different spirit. It is the theory you get when you let yourself be guided by the converse of the Fodorian maxim. Pride of metaphysical place is accorded to qualia. Qualia are seen as the basic reality, the only part of reality that can be experienced directly and of which one can have more than the merely structural knowledge afforded by inference and description.

I opened this discussion with the claim that the task of locating qualia in the world requires a radical reassessment of the standing of percepts and their qualia in the metaphysical order of things. With this reassessment before us we can now specify the physical as well as the metaphysical locus of qualia. The fact that the brain is in the head seemed to show that qualia could not be there. For the brain is made of matter, and grainy matter—so the argument from atomism seemed to show—does not agree with qualia. This way of stating the conclusion of the argument does, however, slightly overstate its force. What the argument from atomism shows is that a quale cannot inhere in matter as one of its properties. But this minor correction does not weaken the case against locating qualia in the brain as long as we do not have an alternative account of how they might inhere in it. Now we have seen another way in which qualia might inhere in matter—the account presented by the other identity theory. On this model, qualia are not properties of matter; they are what matter is made of. This way of bringing matter and qualia together is compatible with the revised conclusion of the argument from atomism. Thus we have reached an answer to the second problem (see page 292) confronting the monist's attempt to house qualia in the brain. The first problem that faced the monist's proposal—the colored-brain problem—was resolved in Section 10.6.2.1. Hence we are finally in a position to state the monistic solution to the nature-problem. Qualia are spatially located in the brain where, metaphysically speaking, they are the basic reality designated by the concepts of neurophysiology.

10.6.3.4 Events, Matter, and Grain

Even the most ardent champion of the other identity theory will have to admit that it cries out for further specification. What we have so far is merely a pointer in the general direction in which a satisfactory account of the relationship between qualia-bearing percepts and neural processes might be found. I believe that Russell's event ontology provides a framework that allows one to recast the basic ideas of the other identity theory in a fashion that sheds a little more light

on the details of the theory. After sketching the barest outline of Russell's event ontology, I want to apply it to two sore points of the other identity theory. First, I want to raise a question related to the colored-brain problem. Second, I want to come back to the grain problem and offer a somewhat more detailed account of the solution to this difficulty. Russell's event ontology will make one final appearance in the next section where it will help to lay to rest the specter of panpsychism.

Russell argues that modern physics is best accommodated within an event ontology. Events, not permanent substances, are the stuff of the world—they are the elements from which mind and matter are to be logically constructed. An event is "something occupying a small finite amount of space-time." (Russell 1979: 222) Examples of events are "an explosion, a flash of lightning, the starting of a light-wave from an atom, the arrival of the light-wave at some other body" (Russell 1979: 87) and "seeing a flash of lightning…hearing a tyre burst, or smelling a rotten egg, or feeling the coldness of a frog." (Russell 1979: 222) Some of these events are percepts—these are the only events whose intrinsic nature we know; others are inferred from percepts—only their structural features are known. The brain, like all other matter, is made up of physically primitive units which in their turn are constructed from events that occur in that region. Among the events that occur in that region are percepts. So some of the events that compose the basic physical particles that compose the brain will be percepts.[157] Stated without qualifications, the resulting view says that "the events that make a living brain are actually identical with those that make the corresponding mind." (Russell 1956a: 147) Though this minimal exposition only scratches the surface of Russell's elaborate account, it can serve to advance the understanding of the other identity theory.

Feigl's version of the other identity theory tells us that qualia are the reality to which neurophysiological concepts apply. This does secure a place for qualia in the world. But it tells us nothing about the relationship between qualia and the physical organ in the head. The picture conjured up in one's imagination is one of qualia floating lazily in the cranium like goldfish in a bowl. How these aimlessly drifting qualia are supposed to jell into the neuronal structures the physiologist observes when he peers into your head we do not know. The monist's structural realism commits him to the existence and observability of these structures. So we need an account that bridges the gap between the basic reality (the qualia) and the higher-level structure (the brain) in your head. Feigl gives us no hint as to how these incongruous levels are supposed to meet. But the Russellian story suggests an answer to this question. One's qualia make up the elementary particles, which make up the atoms, which make up the molecules…which make up the brain. So the qualia enter at the very bottom of an elaborate series of

structures the highest level of which is the organ in our head. Since the brain is made of elementary particles and the groups of events that we call 'elementary particles' are composed (in part or in whole) of events that are percepts, percepts can rightly be viewed as the basic reality. This still leaves us with the complex question how atoms are to be constructed from events. The example cited below provides a rough idea of how this construction is to be carried out. But the introduction of the chain of structures stretching from individual events to global brain structure sheds some light on the question with which we started, even if the first link of the chain continues to harbor some obscurity.

The second issue that is profitably addressed from the perspective of Russell's construction is the grain problem as it presents itself in the argument from atomism. The attempt I made to save the other identity theory from the grain problem (see page 301) may seem less than satisfactory. While the point may be technically correct, it does not effectively alleviate the puzzlement to which the grain problem gives rise. One still does not understand how the simple, homogeneous qualia manage to be in the grainy brain. Russell's event ontology suggests a somewhat more illuminating answer to this question. Since the events that compose the small, discrete particles that give the brain its grainy texture need not share the properties of the objects they compose, they can be relatively large and homogeneous. Thus the simplicity and homogeneity of a felt quale-the red expanse afforded you by looking at the ripe tomato—is properly located at the level of the events that compose the physical particles that compose your brain. The red expanse is extended not in the physical space which we infer from our percepts; it is extended in the subjective space of your experience. That is to say, it takes up a certain portion of the space of your visual field. This extended red expanse and other events like it form the raw material from which we can construct the physical particles whose existence we infer based on the percepts we have when we look at your brain. This gives a bit more substance to the idea that problems of grain arise from a confusion of levels. Simplicity and homogeneity belong to the events out of which complex and inhomogeneous material particles are constructed. If these distinct levels are kept apart, the grain problem doe not pose itself.

The plausibility of this amended answer to the grain problem will turn, to a large part, on the nature and plausibility of the proposed constructions. A simple example that Russell uses to illustrate the method is the construction of an experienced moment out of temporally extended percepts:

> We assume that two events may have a relation which I will call "compresence," which means, practically, that they overlap in space-time. Take, for instance, notes played by different instruments in orchestral music: if one is heard beginning before the other has ceased to be heard, the audi-

> tory percepts of the hearer have "compresence." If a group of events in one biography [roughly, the group of events forming one's life] are all compresent with each other, there will be some place in space-time which is occupied by all of them. This place will be a "point" if there is no event outside the group which is compresent with all of them. (Russell 1954: 294)

It is not hard to see how to generalize this example into the spatial domain. Making the necessary changes, we can use it to construct a point on a line. This brings the example somewhat closer to the case we are interested in. It gives us some idea of how might use a group of extended events to construct something that has the features of a physical point-particle. In this way the other identity theory enriched with the apparatus of Russell's event ontology sheds a bit more light on the question how homogeneity and grain might be combined in a non-mysterious manner.[158]

10.6.4 The Threat of Panpsychism

The monist introduced the other identity theory because it promised to reserve a place in the world for the monist's percepts with their self-presenting, silver-screen qualia. While the critic may grant that the other identity theory has kept this promise he will add that it has kept this promise in a way that should make the monist very unhappy. For instead of reserving a small but exclusive private back room for the monist's qualia, it placed a reservation on the entire physical universe. Now the qualia are everywhere: panpsychism threatens; and it is, as Fodor might say, "the end of the world". (Fodor 1989: 77) The way in which I have presented the other identity theory—speaking loosely about the identification of matter with qualia—may have encouraged this worry.

The advocates of the other identity theory are aware of this worry and anxious to set the record straight. Feigl, for example, explicitly contrasts his version of the identity theory with panpsychism:

> In the physical account of the universe as provided in the four-dimensional Minkowski diagram, there are sporadically some very small regions (representing the brains of living and awake organisms) which are "illuminated by the inner light" of direct experience or sentience. This view differs from panpsychism which assumes that the 'internal illumination' pervades *all* of physical reality. (Feigl 1958: 451)

Feigl argues that panpsychism rests on a faulty inductive generalization. Our acquaintance with the referents of physical descriptions is limited to a small number of highly specific neurophysiological terms. Hence, there is no reason to accept the hypothesis that the qualitative content of the referents of all the other physical descriptions (with which we lack acquaintance) is the same or even

similar to the content of our brain states. The analogy between certain of our brain states and the rest of the physical states in the universe is too weak to underwrite the leap to panpsychism. This is the idea that Feigl summarizes in the following passage:

> Speaking "ontologically" for the moment, the identity theory regards sentience (qualities experienced, and in human beings knowable by acquaintance) and other qualities (unexperienced and knowable only by description) the basic reality. In avoiding the unwarranted panpsychistic generalization, it steers clear of a highly dubious sort of inductive metaphysics. (Feigl 1958: 474)

Feigl sees the universe as a web of qualities. The small number of qualities we experience directly are mental. We remain ignorant of the nature of the vast majority of qualities, and there are no good reasons for thinking that these qualities are mental too. The accusation of panpsychism is therefore unfounded. 'Pan-quality-ism' would be a more accurate label of the view. (See Feigl (1970: 40))

These same considerations figure prominently in Russell's thought on the matter. While Feigl never seems to question the intrinsic mentality of the qualia, Russell takes the issue a significant step further. According to Russell, the percepts (with their qualia) are neutral as between the mental and the physical. The neutral theory of experience with its fusion of act and object strongly suggests the neutrality of the resulting percept. There is, Russell argues, nothing intrinsically mental about percepts. We cannot detect in them a special feature that sets them apart from all other events if only for the fact that "we do not know enough of the intrinsic character of events outside of us to say whether it does or does not differ from that of 'mental' events." (Russell 1979: 171) Thus, we get the conclusion that "the stuff of the world may be called physical or mental or both or neither, as we please; in fact, the words serve no purpose." (Russell 1979: 112) Mind and matter are late arrivals on the scene—both are constructions out of intrinsically neutral elements. This is the cornerstone of the doctrine of neutral monism. By adopting neutral monism one rises above (or rather, drops below) the materialism/panpsychism dispute. In this way the Russellian version of the other identity theory provides another response to the accusation of panpsychism.

The metaphysics of neutral monism has many virtues and is in keeping with some of the assumptions guiding this investigation (see, e.g., the neutrality thesis espoused in Section 2.2.2.3 and its elaboration in Section 7.6.3). However, its claim to lay to rest the specter of panpsychism is not its strongest point. Neutral monists have a deplorable tendency to suspect each other of idealistic or panpsychistic leanings. Russell, for example, chided both Mach and James for choosing 'neutral' elements that leaned too much toward the mental side. And

those who are not neutral monists themselves tend to suspect *all* neutral monists of being idealists, or panpsychists, or phenomenalists, or whatnot, thinly disguised under a blanket of words.

Though I find Feigl's and Russell's rebuttals to the charge of panpsychism convincing, I do not expect all the critics to be silenced. So the question arises what the monist is to do when confronted with a panpsychism charge that refuses to go away. When faced with this situation I would bite the bullet, having first convinced myself that it is made of rubber—unpleasant but not deadly. What is it about panpsychism that appears as so threatening? First, there is the worry that psyched up matter will refuse to obey the laws of physics. Second, there is the worry that making the heart of matter warm and fuzzy is simply perverse. The other identity theory speaks to both of these concerns and shows them to be less threatening than they appear.

All the proponents of the other identity theory argue that making qualia into the basic reality of the brain does not compromise the truth of physics. They see materialism as a cluster of theses among which is the crucial thesis of the universal rule of physical law. They are concerned to preserve this part of materialism, while rejecting what they see as a confused "imaginative picture of matter." (Russell 1954: 382) In this way, the traditional doctrine of materialism is exploded while its worthwhile component is preserved. Here is how the different strands of materialism come apart in Feigl's hands:

> I am indeed in agreement with one main line of traditional materialism in that I assume...that the basic *laws* of the universe are the *physical* ones. But...this does not commit me in the least as to the nature of the *reality* whose regularities are formulated in the physical laws. (Feigl 1970: 40)

Russell is in broad agreement with this assessment of the role of physical law.[159] The reconceptualization of matter in terms of groups of neutral events makes materialism obsolete but

> those who would formerly have been materialists can still adopt a philosophy which comes to much the same thing in many respects. They can say that the type of causation dealt with in physics is fundamental, and that all events are subject to physical law. (Russell 1979: 126-127)

That is, all of your percepts are physical events in the sense that they are members of groups of events (atoms, etc.) whose behavior is described by the laws of physics. In this respect your percepts are no different from the events that you do not experience. Thus, monism poses no threat to an orderly, law-governed world picture. If the critic of monism insists on counting it as a form of panpsychism, we reply that the resulting panpsychism is watered down to a point where

it is too weak to cause anybody to lose their anxiously guarded naturalistic sobriety.

The monist will approach the second worry about panpsychism by recalling the doctrine of structural realism and the considerations that led up to it. The overpowering intuition that the monistic description of matter must be wrong draws its strength from the fact that "it is extraordinarily difficult to divest ourselves of the belief that the physical world is the world we perceive by sight and touch." (Russell 1979: 112) Once one succeeds in thinking of matter as "a mere ghost haunting the scenes of its former splendours" (Russell 1979: 235) the rug is pulled out from under the intuition that nothing bearing even the remotest resemblance to qualia could form the basic building blocks of matter. Here, as in many other cases, it can be enormously liberating to remind oneself of the depth of one's ignorance. The bafflement occasioned by the monistic account of the brain (and the rest of the material world) decreases as our appreciation of our ignorance about matter increases. All of this presupposes, of course, that the structural realism to which the monists help themselves is sound. But we should count it as progress if the complaint about the covert panpsychism of monism were to change into a serious discussion about the picture of matter painted by structural realism.

10.7 Summary

This book opened with the statement that the goal of this investigation was to transform a slogan—to be conscious is to have qualia—into a substantial philosophical thesis. Throughout most of these pages I pursued the relatively safe course of saying what I do not believe. This did not add much substance to the thesis. But fencing off many of the more popular routes to consciousness helped to narrow the search for promising materials with which to firm up the thesis. Thus the task as it presented itself at the beginning of this chapter was to present an account of what it is to have qualia that respects a rather daunting list of constraints. These constraints flow from two maxims gleaned form our survey of the prevailing theories of consciousness.

1. *The theory must not make the act/object assumption.* Analyzing experience on the act/object model places qualia on the side of the object and the act becomes the means by which one gets to have them. Both ends of this story are problematic. Focusing on the act we can summarize what we have learned by saying that none of the proposed act candidates can do their job.

Turning to the object side of the act/object model leads to the second maxim:

2. *The theory must not make the substance/attribute assumption.* Analyzing qualia as properties borne by underlying substances requires the existence of suitable phenomenal objects. We can summarize what we have learned by saying that none of the proposed object candidates can do their job.

Spelling out the problems mentioned in (1) and (2) in greater detail yields four partially overlapping constraints on the nature of qualia and the having-relation to which a theory of consciousness may help itself.

3. *The qualia must not be Mental*—qualia must not be the sorts of properties that need Mental objects (like sense-data) to exemplify them. This constraint is an assumption for which I presented virtually no arguments. It is authorized by the prevailing *Zeitgeist* dominating the philosophy of mind.
4. *The qualia must not be relocated*—qualia must not be the sorts of properties that need either neural states (mental paint) or the surfaces of external physical objects (physical paint) to exemplify them. Grain problems prevent silver-screen qualia from playing either of these roles. The much more pliable celluloid qualia may well be available in both physical-paint and mental-paint versions. But the use of relocated celluloid qualia is excluded because it makes the having-problem intractable.
5. *The qualia must not be of the celluloid variety*—qualia must not be the sorts of qualities whose instances need to be represented in order for them to be felt. The representational accounts of how celluloid qualia are had are unilluminating. They do not explain why representing a celluloid quale results in its being like something to do so. Since there are no nonrepresentational accounts of how celluloid qualia are had, celluloid qualia are off limits.
6. *Having is not representing*—having a quale must not be a matter of believing, knowing, introspecting, or otherwise representing that quale. If the quale is of the celluloid variety, representing it will not make it felt (see (5) above). If the quale is of the silver-screen variety, representing it is not going to impart a feel to it because the quale was already felt in the first place.

I distinguished two theories of consciousness that seemed to satisfy all of these constraints—adverbialism and monism. Adverbialism is the simpler and the more elegant of the two theories. However, it was rejected because it does not satisfy what might be called the supreme constraint on any theory of consciousness: the constraint of phenomenological adequacy. Monism tries to account for consciousness in terms of unborne, self-presenting qualia of the silver-screen variety. Making the qualia unborne does justice to (2) and satisfies (3) and (4). By making the qualia self-presenting one makes them into silver-screen qualia that need not be represented in order to be had. This does justice to (1) and satisfies (5) and (6). And I have argued that the account of consciousness that mo-

nism fashions from these ingredients is phenomenologically adequate. So all is well with monism.

But before celebrating the success of monism we have to attend to the darker side of the doctrine. It may seem that the 'success' of monism is due to its liberal employment of thoroughly miraculous ingredients. How is a naturalistically inclined philosopher to make sense of unborne, silver-screen qualia that present themselves to the subject? I have spent the better part of this chapter sketching out a philosophical framework within which these ingredients of monism begin to make sense. Monism turns out to be a complex patchwork of theories. As a result monism lacks the elegance and simplicity of adverbialism. Let us review some of the main ingredients of the monistic theory of consciousness:

M1. *The neutral theory of experience.* It allows us to overcome the act/object assumption. By identifying act and object we end up with the notion of a percept. Percepts 'accommodate' qualia and present themselves (and their qualia) to the self they (partially) constitute. They are the material from which selves and brains are built.

M2. *The bundle theory of the self.* It is the crucial ingredient of the explanation of how self-presenting qualia are had. The transformation of the person into a bundle of percepts obviates the need for an additional representational act by means of which the person appropriates the quale that inheres in the percept. Thus, qualia can be said to present themselves to the person they constitute. Since these qualia need not be represented in order to be had, they are silver-screen qualia.

M3. *The bundle theory of percepts.* It is the crucial ingredient of the explanation of unborne qualia. Qualia inhere in percepts not as properties in a substratum but as elements in a bundle. Causal considerations place percepts into the head of the perceiver. Thus, what one perceives is in one's brain. Extracranial reality is never perceived directly. This makes it possible to solve the colored-brain problem and it is part of the motivation for structural realism.

M4. *The other identity theory.* It is the metaphysical framework that lets one see how the unborne silver-screen qualia, the percept bundles they form, and the subjects to which they present themselves can all find a place in an orderly world that is governed by the laws of physics. Structural realism is one of the key ideas of this theory. All matter (except one's own brain) is known only as an inferred structure whose nature remains unknown. One's percepts are the only points at which matter reveals its nature to the experiencing subject. The message of the revelation is that brains are, at bottom, qualitative. Qualia are what the brain is. The qualities at the heart of all other matter are unknown. We have no good reason to think that they bear much

similarity to the qualia of our experience. The world of the other identity theory is a distant relative of our common-sense conception of the world. But that this world is odd does not show that it is mysterious or magical in some naturalistically objectionable sense.

On some of these topics more has been said in the present chapter. On all of them much more should have been said. But this sketch of a sketch provides enough information to enable us to state and understand what the monist has in mind when he says that to be conscious is to have qualia. Let us use a concrete example—the ripe tomato, of course—to summarize the monist's meaning when he utters this harmless looking slogan.

The ripe tomato looks red and round and bulgy to you. The redness and the roundness and the bulginess are the qualia that form your tomato percept. In enjoying these qualia you are face to face with reality as it is in itself. Qualia are not appearances of an underlying reality. They are the only part of reality whose nature is revealed to you without inference. The tomato percept is what you truly see—the ripe tomato is an unobservable object whose existence and structural features you can infer from your percept. The percept is located in your brain. Looking at the ripe tomato, therefore, teaches you something about your brain, viz., that the real nature of the matter forming your brain is qualitative. Percepts and their qualia are the reality to which the concepts employed by brain scientists apply. In this way qualia come to occupy a quite central place in the world: they are what your brain is made of. Percepts and their qualia are also what you are made of. You are a bundle of percepts (among other things) and the tomato percept is a member of this bundle. This is what it is for the tomato percept to be present to you. Looking at the ripe tomato is like something for you because it causes qualia of redness, roundness, and bulginess to form a percept that becomes a part of what you are.

Summarizing these monistic answers to the nature-problem and the having-problem, we can say this. Qualia are physically located in the brain where, metaphysically speaking, they are the basic reality that neurophysiology describes. Percepts are bundles of qualia. The self is a bundle of percepts. Having qualia (i.e., being conscious) consists in there being bundles of the first kind among the elements of the bundles of the second kind. Taken together these two answers encapsulate the monistic account of consciousness.

I began by saying I wanted to transform a slogan—to be conscious is to have qualia—into a substantial philosophical thesis. That I have done. Of course I wanted to do much more. I wanted to transform it into a substantial and true philosophical thesis. I do not know whether I have done that. But this is the best I could do.

Notes

1. This is how Herbert Feigl (1967: 138) reports (and translates) Einstein's thoughts about qualitative consciousness. The report is not quite accurate. For Feigl tells us that Einstein used a "rather uncouth" word in the original German. Since nobody can fail to know the word (or, at least, the meaning of the word) that Einstein must have used I need not mention it here.
2. This program is carried out in paradigmatic fashion by Alvin Goldman (1986).
3. Capitalization of the first letter in the words "mind" and "mental" indicates that these words stand for nonphysical entities, events, or properties. The lowercased versions of these words are free of such ontological significance. Context will make clear whether they stand for something Mental or for a physical/functional item that is also mental.
4. The term "thought" is here used in a very broad sense so that things like the red expanse before Aune's mind are included.
5. Longer chapter summaries are found at the end of each chapter.
6. For a sampling of other philosophers who employ this notion of consciousness consult the quotes in Section 2.2.2.7.
7. B.A. Farell (1950) was the first to characterize consciousness by means of the "what it's like" locution. I present Nagel's version here because it was Nagel's paper that is responsible for transforming this unusual formulation into a philosophical catchword.
8. See, e.g., Hofstadter and Dennett (1981: 403-414), Armstrong (1984: 45-66), and Lycan (1987: 76-77, 142-143).
9. See also footnote 20 and page 34. For an interesting development of a position that results from a combination of the first two misinterpretations of Nagel's locution—i.e., from reading the epistemized version in the comparative way—see Carruthers (1992), especially pages 47 and 48.
10. For more on the idea of the "subject involvement" of qualitative consciousness see Section 5.2.4.
11. This is as it should be. Illuminating characterizations of primitive and intuitive notions are hard to come by. One should not expect them to change in 20 years or, for that matter, in 200 years.
12. The most recent and most penetrating presentation of the notion of qualitative (phenomenal) consciousness can be found in Block (1995) and the accompanying commentary by various authors.
13. It seems quite possible that we should have considerable knowledge about the causal conditions that normally give rise to consciousness without much of an understanding of why this should be so or what, exactly, the produced phenomenon of consciousness is. See footnote 87 and Section 10.5.5 for further comments on this question.

14. The neutrality thesis is nothing new. It has been emphasized by H.H. Price, for example. Unfortunately, Price phrases the neutrality thesis in the language of sense-data; but as best I can tell his notion of a sense-datum is not that discussed and criticized in Section 2.2.2.4. Price tells us that when he sees a tomato the following stands fast for him: "that there exists a red patch of a round and somewhat bulgy shape, standing out from a background of other color-patches, and having a certain visual depth, and that this whole field of color is directly present to my consciousness." (Price 1950: 3) Data of this kind Price calls "sense-data." These sense-data are metaphysically neutral because by admitting them "we are not committed to any view about what is called 'the status' of sense-data in the Universe, either as regards the category they fall under, or as regards their relations with other types of existent entities. They may be events, or substances, or states of substances. They may be physical; i.e. they may be parts of, or events in, material objects such as chairs and tables or (in another theory) brains. They may be mental, as Berkeley and many others have held. They may be neither mental nor physical." (Price 1950: 18-19)
15. A good part of the distaste for sense-data stems from a misapprehension of their nature, viz., the belief that sense-data must be construed as nonphysical things. It is worth pointing out that such original champions of sense-data as Russell and Moore took very seriously the possibility of physical sense-data. Giving up on the nonphysical nature of sense-data does not solve all problems. But it should soothe the metaphysical horror that sense-data so commonly instill. See also Section 7.5.1.
16. For a very self-assured rebuttal of Dennett's qualia eliminativism along these lines see Owen Flanagan's (1992). As he sees it, Dennett's attack on qualia simply misses the mark: "There are no qualia with these properties [i.e., the properties that supposedly make up the "fourfold essence" of qualia]. Quine the properties, but don't quine qualia. Leave the real qualia alone." (Flanagan 1992: 65) I share Flanagan's suspicion that the qualia Dennett eliminates are something of a straw man. But regrettably I am even less confident that Flanagan's "genuine qualia" are the real thing. According to Flanagan, a "quale is a mental event or state that has, among its properties, the property that there is something it is like to be in it." (Flanagan 1992: 64) I disagree. Qualia aren't mental events or states. They might have turned out to be properties of mental events or states. But, if I am right, even that is false. So Flanagan and I disagree about the category into which qualia belong. But the disagreement does not stop here. Following Paul Churchland, Flanagan holds that "sensory qualia...are just the characteristic spiking frequencies of activation patterns in the relevant neural pathways." (Flanagan 1992: 54) Here, too, I disagree. In Chapter Nine I explain why I think that spiking frequencies and suchlike aren't suited to be, or exemplify, the sort of qualia from which consciousness is made.
17. For another attempt to come to terms with Dennett's qualia eliminativism see Section 7.2. There I suggest that Dennett is best interpreted not as eliminating qualia but as recategorizing them as ordinary physical properties of external physical objects.
18. This primacy of qualitative consciousness is defended in Section 3.3.2.4. Sections 9.3.4.1 and 10.5.6 are also relevant to this discussion.

19. I make this move inspired by Terence Parsons's provocative demolition of the adage "no entity without identity." Here is the gist of Parsons's reflections: "The importance of the identity maxims in recent literature has been their role in finding easy objections to philosophical theories. Do we have trouble individuating propositions? Then junk them and theory that appeals to them! Likewise for Fregean senses, properties, possible though nonactual objects, etc. Frege, Meinong, and a host of others are seen to have been dabbling in incoherence, and their theories are to be rejected. Clearly this is too facile to be persuasive....A theory may provide us with a great deal of understanding of the world, in spite of its not giving us completely clear answers to identity questions in general. Whether this is so in a given case depends on what the theory says, and cannot be decided in a simple *a priori* fashion, based on simple maxims." (Parsons 1987: 14-15) I hope to show that qualia are sufficiently important for our understanding of consciousness to earn their ontological keep, problems concerning their identity notwithstanding.
20. *You* would believe that I am fine. But you would be terribly wrong. These considerations barely hint at the enormous epistemic problems that qualia raise, both concerning our knowledge about our own qualia and concerning our knowledge about the qualia of others. Throughout this book I will resolutely avoid epistemological questions of this sort. These questions are interesting and sooner or later we must attempt to answer them. I opt for later, guided by the belief that an understanding of what qualia are and how they are had must precede an account of what it takes to know them. This is a momentous tactical choice. If it is mistaken—a possibility that I cannot exclude—then much of the work done here will be pointless.
21. The consideration of belief feelings raises the difficult question how the believed content (as opposed to the attitude) is presented to the mind. How does one present to oneself that which one believes? Could a mind devoid of all sensuous phenomenology still direct itself upon a specific proposition? Russell provides a thought provoking discussion of these issues in Chapters 11 and 12 of his *Analysis of Mind.* As I understand it, it is a consequence of Russell's view that a mind devoid of all sensory qualia could not relate itself to a content in any form; it could not believe, remember, or think anything. The discussion of this problem is still going strong at this time. Those who try to spell out in a psychologically credible way the modes of presentation with which we present propositions to ourselves deal with the same problem. Alvin Goldman's (1993b) provides a fascinating contemporary perspective on some of these questions.
22. No one has driven home this point more insistently than Sellars. This simple but disturbing observation forms the intuitive heart of his notorious grain problem. For a discussion see Sections 9.3 and 10.6.3.4.
23. One misunderstanding should be blocked right here. Surely the bafflement is not grounded in the untenable doctrine that no two "very different" properties can be coinstantiated. Nevertheless, from the admission that some very different properties can be coinstantiated, it does not follow that any two properties can be coinstantiated. This is obvious for the pair *red all over* and *not red all over*. It is, perhaps, a little less obvious for the pair *red all over* and *green all over*. And it may not seem at all obvious why the pair *being a brain*

process and *being phenomenally red* should be incompossible properties. I discuss these issues in Chapter Nine.

24. For more on the colored-brain problem see Sections 7.5.2, 8.3 and 10.6.2.1.

25. The strategy of tackling philosophical problems by doing semantic violence to the word "in" has been used often—"there is remarkably little, not only in the history of idealism, but in the entire history of philosophy, which does *not* hang on the world 'in'", as the iconoclastic David Stove (1991: 157) has said—but succeeded rarely. This does not bode well for my own attempt to come to terms with the problem of unobservability. But I do not know how to avoid this unpromising course.

26. The wavering between the terms "materialism" and "materialistic science" conceals a deep issue. It can be argued that science as such is not committed to materialism (or any other metaphysics, for that matter). And it can further be argued that adding a materialistic metaphysical underpinning to science does not increase its explanatory potential. Therefore, the explanatory fecundity of "materialistic science" is due only to science, not to materialism. Giving up materialism does, therefore, not amount to much as long as one retains the science. I have tried to spell out his thought in a bit more detail in Stubenberg (1996).

27. Compare Russell's related comments about the relationship of the physical and the real: "This belief in the physical world has established a sort of reign of terror. You have got to treat with disrespect whatever does not fit into the physical world. But that is really very unfair to the things that do not fit in. They are just as much there as the things that do. The physical world is a sort of governing aristocracy, which has somehow managed to cause everything else to be treated with disrespect. That sort of attitude is unworthy of a philosopher. We should treat with exactly equal respect the things that do not fit in with the physical world, and images are among them." (Russell 1918/19: 257)

28. Michael Lockwood describes his position (that is very similar, at least in inspiration, to the one defended here) as materialism without physicalistic reductionism. (See Lockwood (1993: 272)) Howard Robinson seems to agree, if somewhat grudgingly, when he says: "Although apparently bizarre, this is the nearest to a plausible form of materialism there is." (Robinson 1993: 23) A brief comment about the relationship between materialism and the view I defend in Chapter Ten is contained in Section 10.6.4.

29. For an exposition and critique of Armstrong's view consult Section 9.3. My selecting Armstrong as the color realist who has to take all the heat about phenomenological adequacy may convey the mistaken impression that Armstrong has an especially prominent blind spot concerning phenomenology. Nothing could be further from the truth. No other color realist equals his sensitivity to the problems raised by phenomenology. Here, for example, is his most recent "confession": "I confess that I myself am still not fully satisfied with the physicalist reduction [of color]. I accept the reduction; I advocate it. But the phenomenology of the affair continues to worry me." (Armstrong 1987: 12) That a philosopher so sensitive to issues of phenomenology can, in the end, give up the condition of phenomenological adequacy worries me. Perhaps it indicates a well developed blind spot of some sort on my own part. See also Armstrong (1979), where there is a similar acknowl-

edgment of phenomenological inadequacy, together with the maxim that theoretical concerns overrule phenomenological concerns.

30. My repeatedly mentioning introspectionist psychology should not be misconstrued as an endorsement of this position. As Chapter Five and Chapter Six will make clear, I hold that the attempt to understand consciousness in terms of introspection is mistaken.
31. Dennett's expression "materialistic perspective of contemporary science" is conveniently vague. I trust that Dennett does not require the friends of the first-person perspective to be strict materialists. For if this were what it takes to avoid obscurantism, Dennett himself would reside on the dark side. He would be the first to admit that his version of functionalism is a far cry from hard-core materialism.
32. In "Three Kinds of Intentional Psychology" Dennett (following Reichenbach) distinguishes between "two sorts of referents for theoretical terms: illata—posited theoretical entities—and abstracta—calculationbound entities or logical constructs." (Dennett 1987c: 53)
33. For the autophenomenologist *esse est percipi* (or something close to it) is true of sensations. The reality of a sensation consists in its being had by someone. The question whether it is identical with a neurophysiological process is an interesting question concerning the nature of sensations awaiting further discovery. But no such discovery could show that a sensation that is currently being had by someone does not exist.
34. Notice that the autophenomenologist's questions would not have been answered, even if the heterophenomenologist had succeed in "trading in" his posits for some neural processes. For we would still be in the dark as to how this could be so.
35. Here is how Dennett has characterized this uneasy relationship some years ago: "Philosophers of mind and epistemologists have much to learn from recent work in cognitive psychology, but one of philosophy's favorite facets of mentality has received scant attention from cognitive psychologists, and that is consciousness itself: full-blown, introspective, inner-world, phenomenological consciousness. In fact if one looks in the obvious places...one finds not so much a lack of interest as a deliberate and adroit avoidance of the issue. I think I know why. 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 a respectable theory." (Dennett 1978: 149) Things have slightly improved in the meantime. In a more recent assessment of the strained relationship Dennett writes: "Consciousness is making a comeback in psychology, but there is still residual skepticism, anxiety, and confusion about how to approach this perilous phenomenon scientifically." (Dennett 1982: 159) For two more recent views of the role of consciousness in the contemporary cognitive sciences, see A.J. Marcel's and E. Bisiach's (1988: 1-15) interesting introduction to the volume *Consciousness in Contemporary Science* and Flanagan (1991: 307-14). For a promising attempt to secure a place for phenomenal consciousness in scientific psychology, see Goldman's (1993a; 1993b) account of the mechanism underlying our ability to self-attribute psychological states. Of course, the case for the scientific recognition of con-

sciousness is not helped by the fact that some of the leading qualophiles insist on the epiphenomenal character of consciousness.

36. Just in case you think that only the unduly soft or tender minded could feel that way, you may want to consider what David Armstrong—Australian arch materialist and definitely *not* the epitome of tender mindedness—has to say on this matter. Speaking of the sort of consciousness that "seems like a light switched on, which illuminates utter darkness" (Armstrong 1978: 63) Armstrong says that without it "we would not be aware that we existed—our self would not be self to itself." (Armstrong 1978: 67)
37. Context permitting, I will use the labels "HOP," "HOT," and "HOR" without always adding the words "theory of consciousness."
38. A comment on terminology. Rosenthal calls this high-grade consciousness "introspective consciousness." Lycan uses this term to stand for ordinary consciousness which Rosenthal calls "nonintrospective consciousness." When Lycan means to talk about this high-grade consciousness he talks of consciousness that results from introspect*ing* (i.e., the activity of introspecting deliberately).
39. For another attempt to critically employ Pollock's characterization of blindsight see Sections 6.4.3-6.4.6.
40. Pollock uses the term "sensation" to stand for the outputs of a perceptual sensors. In the absence of introspective sensors these outputs will not be endowed with phenomenal properties. That is why Pollock can consistently characterize a blindsighted person as having visual sensations.
41. To my amazement I found that not quite all agree. Here, for example, is a characterization of the driver's consciousness that makes it look suspiciously like the sort of consciousness enjoyed in blindsight or (presumably dreamless) sleep: "It isn't like anything to be an automatic, long-distance driver. Or, to play a bit with 'like' idioms, being an automatic driver is like being asleep; and it isn't like anything to be asleep." (Sanford 1984: 70)
42. This criterion/condition may strike you as preposterous. You may wonder what this sort of reactivity has to do with consciousness. Does it mean that you should count the friendly door at the local supermarket as a member of the charmed circle on the grounds that it opens politely whenever it "sees" you coming? For more on the "reactivity criterion" of perceptual consciousness see Section 4.3.1.2.
43. But see the discussion of Helen in Section 6.4.5.
44. As do Armstrong and Pollock—the other two HOR theorists we encountered—in their own ways. Pollock's account of the origin of qualitative consciousness is the focus of Chapter Six. And to Armstrong we owe the powerful image of introspective consciousness illuminating what would otherwise be "utter darkness." (See p. 79).
45. Here, for example, is how Rosenthal puts the matter: "It is crucial to avoid a merely verbal issue. Some find it tempting to hold that the term 'sensory quality' can apply only to those qualities by reference to which we say what it is like to have one or another conscious sensation. If so, nonconscious states plainly cannot have sensory quality. Similarly, sensory states might be held to be definitionally conscious states." (Rosenthal 1991: 21)
46. Although the hierarchical model of introspection is dominant, it is not the only one. See Lyons (1986) for a discussion of a number of competing theo-

ries of introspection. See also Lycan (1987: Chapter 6), and especially Hill (1991), for a defense of the "volume control" model of introspection. For a critique of the hierarchical model of consciousness, see Lyons (1986: Chapters 3&4).

47. For more on one central issue raised by this passage see Section 4.4.7.
48. I have discussed his views at considerable length in Stubenberg (1992a).
49. For more on the Armstrongian account of perception, see Section 6.4.1 where I come back to the topic of perception devoid of a phenomenology. For a discussion for a related but slightly different topic see Section 8.4.2.
50. It should be noted that these authors do *not* take that to be reason enough to embrace HOT.
51. Some clarifying comments are in order here. First, by pointing out this deep divide within introspectionism I am not going back on my claim that HOT and HOP should be treated as a unified phenomenon. They are unified inasmuch as the differences between the "T" and the "P" are irrelevant for our purposes. But that is not to say that there cannot arise other disagreements within introspectionism. Second, I have presented HOR (and, by implication, introspectionism) as the theory that addresses the having-problem, not the nature-problem (see Section 4.3). But the relocation issue belongs primarily under the heading of the nature-problem. To this one could react in two ways. One might say that the location question is not, strictly speaking, a question that HOR has to solve. To solve this question is the task of the "metaphysical housecleaning" that has to accompany every more specialized philosophical theorizing. Or one might simply acknowledge that HOR does also address the nature-problem, pace the vehement denials of its champions. I should be happy to amend my assessment of HOR's scope if the second route is taken. For it makes HOR's future look all the more bleak. Third, the distinction between external and internal qualia must not be confused with that between celluloid and silver-screen qualia (see Section 4.3.1). The current debate is an in-house debate within the celluloid camp. A parallel debate could (and has) taken place among the friends of the silver screen.
52. For a notable exception see the interesting exchange in Armstrong (1984b) and Rosenthal (1984).
53. C.D. Broad (1969: 280) and A.J. Ayer (1986: 79-80) seemed to enjoy making this point (but without embracing it). Perhaps Wilfrid Sellars's position comes close to qualifying.
54. I want to forestall one possible misunderstanding of this critique of introspectionism. In arguing that introspection has no role to play in the explanation of qualitative consciousness, I am not trying to insinuate that introspection has no role to play in our cognitive economy as a whole. It is plausible to maintain that by introspection we come to be cognitively related to our qualitative mental states. Introspection is what elevates these states from merely being felt to states that we believe or know ourselves to have. In transforming our noncognitive feelings into epistemically evaluable states, introspection serves to integrate our qualitative states into our cognitive economy.
55. Block's distinction between phenomenal and access consciousness is similar to the one I draw here. In Block's capable hands this distinction becomes a powerful separation method. He shows that access consciousness is the true explanandum of many theories of consciousness that are marketed as theo-

ries of phenomenal consciousness. See especially his Block (1995). David Chalmers aggressively pursues a similar cause using the labels "the easy problem" and "the hard problem." He is relentless in his fight against false advertising in the field of consciousness studies. See his Chalmers (1994) and, especially his Chalmers (1996).

56. Note that in this use the term "consciousness" does not refer to qualitative consciousness but to some kind of cognitive internal access to qualitative consciousness.
57. Not even Berkeley did that. He argued ingeniously that the two should be identified. That is a far cry from simply confusing them.
58. I borrow the term "qualitative belief" from Shoemaker. Shoemaker introduces this term as follows: a qualitative belief is "a belief to the effect that one feels a certain way (or, more abstractly, that one is in a state having a certain qualitative character, or, in still other terms, that one has a certain qualitative state)." (Shoemaker 1975: 189)
59. The epistemological literature discusses this issue under the label "KK-thesis." See, for example, Chisholm (1989: Chapter 10).
60. This distinguishes the qualitative beliefs of the thesis of constant cognition from the introspections of introspectionism. One's failure to notice an introspection postulated by introspectionism proves nothing. For introspections are not transparent.
61. For a contemporary defense of a quasi-Russellian acquaintance principle see Ackerman (1987).
62. Russell even seems to acknowledge (grudgingly) that we cannot exclude the possibility that some mental states are not merely unintrospected but that they are unintrospectable. "It would seem, however, that, when we are acquainted with an object, our acquaintance with it can usually be discovered by an effort of attention—though it must be confessed that, if it could not be so discovered, it is hard to see how we could ever know that it existed undiscovered." (Russell 1984: 121)This point is relevant for the analogy we are setting up. If some mental states are in principle unintrospectable, the analogy between them and tables, for example, becomes much closer than it first might have seemed. There is, of course, also a much more "direct" reason why the existence of such states would be relevant in the context of a critic of the epistemic argument for introspectionism. Such states would falsify premise (2) of the epistemic argument, thereby closing the case against it.
63. Let me just mention two reasons for doubt. Concerning the first question. Let us assume that there are, and that we are in possession of, truths connecting qualitied mental states (with which we are not acquainted) with other mental states with which we are acquainted. Based on such knowledge, you might, for example, infer that you must be suffering a pain now given the fear, the anxiety, and the nervousness you feel. The worry is this: Can't we always imagine that an unqualitied state with the same causal profile as genuine pain is triggering the mental states with which we are acquainted? If so, our inference will be mistaken.

    Concerning the second question. What is involved in having a qualitative belief? Upon being exposed to a red light, a blindsighted and a seeing person may both believe that they are currently in a mental state exemplifying a red quale. And they may both be right. Nevertheless, there appears to be a strik-

ing difference between their beliefs. The seeing person's quale figures into her belief in a way in which the blindsighted person's does not. It seems not unreasonable to hold that qualitative beliefs have to be akin to the seeing person's belief and that the blindsighted person's belief does not qualify.

64. See Lehrer's discussion in Lehrer (1990: 67-68). Lehrer does acknowledge that there exists a very small number of propositions of which we have special knowledge; see *op.cit.* pages 47-48.

65. I join the ranks of the orthodoxy without further discussion of these arguments. As will become evident in the ensuing discussion, my reasons for rejecting the special-knowledge thesis (and the thesis of constant conjunction in general) stem from somewhat different considerations.

66. The terms "identity thesis" and "containment thesis" are Pollock's. See Pollock (1986: 32-33) for a discussion of the simple identity and the containment theses. Note that Pollock discusses these two theses in the context of what I have called the special-knowledge thesis. He argues that neither the containment nor the (simple) identity thesis can serve to support the special-knowledge thesis.

67. I want to comment briefly on the epistemic relations between these various versions of the belief thesis. We do encounter a rather odd epistemic situation here: the *weaker* conditional or biconditional versions of the belief thesis seem to derive whatever plausibility they have from our more or less tacit acceptance of the *stronger* identity or constitution theses. The latter theses seem to offer an *explanation* of the weaker versions of the belief thesis. They help us understand why it is that the conditional or biconditional versions of the belief thesis hold. Were it not for our acceptance of the identity and/or constitution theses, the belief thesis might appear quite unsupported. If this diagnosis of our tendency to accept the belief thesis is correct, we could mount an attack on the belief thesis by attacking the theses that support it. The attractiveness of this strategy consists in the fact that the identity and constitution theses are stronger and therefore more vulnerable than the belief theses themselves. Thus, our chances of successfully criticizing the belief thesis will be high on this approach. The drawback of proceeding in this manner is that, even if it is entirely successful, we will not quite succeed to show what we want to show. For we will not be able to show that the belief thesis (in any of its forms) is false. All we can hope to establish is the conditional conclusion that if the identity and constitution theses are the sole support of the belief thesis, then the belief thesis is devoid of all support. This would not be an uninteresting result, but we can do better than that by mounting a direct attack on the weaker versions of the belief thesis.

68. Compare Block's way of illustrating the notion of phenomenal consciousness in Section 2.2.1.

69. From now on I will use the expression "belief thesis" to refer to the particular interpretation of the belief thesis that reads it as specifying a necessary condition on consciousness.

70. Just a reminder: the appeal to phenomenology is not meant to establish (directly) the existence of unintrospected qualitative states. The introspectionists have a powerful reply to any direct attack based on phenomenological data (see Section 5.3.2). The appeal to phenomenology is directed against the thesis of constant cognition (in the guise of the belief thesis) and

the claim that each qualitative state is accompanied by a corresponding qualitative belief (where this belief is distinct from the introspections posited by introspectionism).

71. For further references see Moser (1989: 80-88, especially fn. 26 on pages 86-87).

72. That is, the first appeal to phenomenology serves to make a point concerning the having-problem: to have qualia (in the relevant sense of this term) is *not* a matter of tokening the corresponding qualitative beliefs. Nothing is said here about the nature of qualia. The second appeal to phenomenology (Section 7.4.1) (unsuccessfully) aims at making a point about the nature-problem: experience teaches that the qualia are properties of the external experienced object, not the mental event of experiencing these objects. Nothing is said here about the having of qualia. There are thus two quite distinct lessons that phenomenology supposedly teaches (or fails to teach). Without a clear understanding of the different thrusts of these two phenomenological appeals we cannot begin to evaluate them properly. For more on how the having of a quale is related to issues concerning its nature and exemplification, see especially Section 7.6.1 and also Section 7.5.3.

73. This, by the way, is very much in keeping with the introspectionist doctrine that most of our introspections are themselves unintrospected and therefore not something of which we are aware. Hence, the view that consciousness arises from introspection is not a view that introspection can discover—see also Section 5.3.2.1.

74. The paper by this wonderful title is cited by Radner and Radner (1989: 7).

75. Compare Dretske (1995). For an exposition (but not an endorsement) of this kind of argument see also Carruthers (1989).

76. As pointed out above, this discussion is focused on hierarchical accounts of introspection. Rival accounts of introspection may allow for the possibility of introspection without metamental processing. Such views would be immune to the version of the argument from simple minds presented here.

77. The neat distinction between the strong and the weak versions of the argument from simple minds may become blurred on certain accounts of introspection. If one were to hold, for example, that introspecting one's qualitative states just is a matter of one's forming the corresponding qualitative beliefs, then the two versions of the argument would collapse into one.

78. There exist other more involved arguments in support of this particular form of skepticism. Peter Carruthers (1989), for example, has recently argued the case against animal consciousness. Based on a Dennettian theory of consciousness, he argues that most (and perhaps all) animals are completely devoid of conscious experience. It seems that he would be prepared to argue with Malebranche that animals "eat without pleasure, cry without pain, grow without knowing it; they desire nothing, fear nothing, know nothing." (Quoted in Radner and Radner 1989: 60) Carruthers's argument, and other arguments like it, proceed from massive philosophical assumptions. In this particular case the arguments rests on Dennett's controversial account of consciousness. Since a discussion of these presuppositions would lead us too far away from the issue of animal consciousness, I will not pursue it here.

79. To learn about the current state of the Oscar Project and to download a current version of Oscar visit his homepage at http://www.u.arizona.edu/~pollock.

I have chosen to discuss Pollock's Oscar because the Oscar project is one of the most recent, most perspicuous, and most straightforward attempts to present an introspectionist account of consciousness.

80. Remember that for Pollock sensations are simply outputs of perceptual sensors. Having a qualitative feel is not part of this notion. Compare also footnote 40.

81. Pollock notes that his account of consciousness is similar to that of Armstrong. I will therefore take the liberty to use some of Armstrong's notions in the ensuing discussion of Pollock's theory.

82. It should come as no surprise that Armstrong himself is not only keenly aware of, but also draws the reader's attention to, the "phenomenological paradox" to which his theory of perception gives rise. See Armstrong (1984c: 17). See also footnote 29.

83. Refer back to Section 4.4.5 to better appreciate the radical nature of the account of perception discussed here.

84. Nicholas Humphrey is the most radical advocate of the independence of sensation and perception. In his Humphrey (1992) he defends the thesis that sensation and perception are "alternative and essentially nonoverlapping ways of interpreting the meaning of an environmental stimulus arriving at the body." (Humphrey 1992: 47)

85. H.D. Mellor, quoted in Natsoulas (1982: 91). For a book-length study of blindsight see Weiskrantz (1986).

86. I have taken this term from Gunderson's rather pessimistic assessment of the prospects of strong AI: "To aim at...an actual incorporation in a strong AI project of program resistant aspects of the mind and the phenomenology they involve seems hopelessly utopian. But any of the more modest alternative models one can imagine seem haplessly prosaic. They all have that *representation-by-fiat-only* ring to them: as if by simply naming some operation or item in a program of flow chart and saying "*This* (affect, etc.) is represented in the program by *that*" and then proceeding merrily to describe input-output patterns, one believes that an illuminating model of affect has taken place. This is all too much like drawing the pineal gland, making a dot in it, and saying "This (pointing to the dot) represents the human soul as it is joined to the human body." Well of course it does, but it tells us nothing instructive about the human mind and its relation to the body, but only something about how easygoing and unilluminating certain forms of representation can be." (Gunderson 1985: 237).

87. One does *not* need to have dualistic leanings in order to appreciate that the problem of *how* a physical system manages to instantiate consciousness might be formidable. Robert Cummins, for example, is keenly aware of this issue: "We could, I think, be in the position of knowing beyond serious doubt that we are conscious because we instantiate a certain functional architecture, yet have no understanding of how this could be. We would, of course, know in a very abstract way what *sort* of phenomenon it is, viz., the sort of thing that supervenes on having a certain functional architecture...But we might still be quite in the dark as to *how* consciousness does or could supervene on that or any functional architecture. We might have no idea, in fact, what sort of intellectual tool it would take to understand this fact. We might not even know whether we should be doing phenomenology, neuro-science,

AI or what." (Cummins 1984: 795-796) But there are also those who deny the existence of the mystery outright. See, for example, Lycan (1987: 50-52, 76-77).

88. Cf. also Maudlin (1989: 401, fn. 1). There Maudlin argues that the specification of conditions that are necessary and sufficient for a system to have conscious mental states constitutes only the smaller part of a satisfactory solution of the mind-body problem. He maintains that "it might be possible to discover such conditions and still be puzzled about how subjective states arise from certain physical conditions. Making the connection between mind and body intelligible, seeing how subjective states can be brought about by objective conditions, is a further problem, and of another order of magnitude."

89. An acute awareness of this mystery has led a number of philosophers to embrace some rather unpopular views. I believe that Russell's neutral monism and Nagel's panpsychism and the current resurgence of dualism can all be viewed as resulting from attempts to resolve this mystery.

90. One might want to object as follows: You do have every reason to assume that you are a typical representative of the Oscar family. For, after all, Oscar and you have the same rational architecture. Thus nothing stands in the way of reasoning analogically from your own case to the case of Oscar. I think that an appeal to the shared rational architecture won't support this inference. The argument from blindsight was intended to show that the functioning of the human rational architecture need not be accompanied by the having or experiencing of qualia. Hence, having this rational architecture is not sufficient for being conscious. In respects other than rational architecture Oscar and I are very different indeed. Since, moreover, we do not know which of my other features (other, that is, than my rational architecture) allow me to have and experience qualia, it would be rash to conclude that Oscar must be similar to me in whatever the relevant respect might be.

91. Given my reading of Dennett's amazing paper "Where Am I?" Dennett (1981) it strongly supports my case against the thought-experimental solution of the question concerning Oscar's consciousness. But I hesitate to make much of this rare instance of apparent agreement because the interpretation of this piece is notoriously difficult.

92. For an earlier attempt to come to terms with Dennett's apparent qualia eliminativism see Section 2.2.2.5. There I portrayed Dennett as an eliminativist. Here I try to make him look like a relocationist. Other passages suggest that Dennett is what we might call a "mental reductionist" about qualia: qualia are nothing but collections of other mental states. This view is suggested when Dennett speaks of "the "reductionist" path of *identifying* "the way it is with me" with "the sum total of all the idiosyncratic reactive dispositions inherent in my nervous system as a result of my being confronted by a certain pattern of stimulation." (Dennett 1991: 387) And when lecturing Otto—the imaginary interlocutor whose role it is to defend the views of the traditional qualophile—Dennett puts the matter even more explicitly: "What qualia *are*, Otto, are just those complexes of dispositions. When you say "*This* is my quale," what you are singling out, or referring to, *whether you realize it or not*, is your idiosyncratic complex of dispositions." (Dennett 1991: 387, 389)

I prefer to think that these shifts in Dennett's view (and of my characterization of it) are less a sign of muddle-headedness than a sign of the fluidity of the boundaries between reduction, elimination, and relocation. Dennett emphasizes this fluidity when he writes: "The difference between 'eliminative materialism'...and a 'reductive' materialism that takes on the burden of identifying the problematic item in terms of the foundational materialistic theory is thus often best seen not so much as doctrinal issue but as a tactical issue: How might we most gracefully or effectively enlighten the confused in this instance?" (Dennett 1988: 44, fn. 2) Another author who makes much of this sort of fluidity is David Lewis (1995).

93. Dennett is a little impatient when dealing with those who read him as an "all out eliminativist" about consciousness Dennett (1991: 45) and Dennett (1993: 45-46). But I think that he did (and still sometimes does) invite this extreme interpretation of his view.
94. We need not now enter into a discussion of Dennett's claim that the physical objects are the home of the secondary qualities. This issue is naturally addressed in the course of a discussion of the wholehearted relocation thesis (see the following Sections and Chapter Nine).
95. Kim makes this comment about a similarly radical interpretation of content externalism. One can only guess what he would have to say about consciousness externalism.
96. Robinson's analysis of what went wrong in Moore's account of consciousness differs somewhat from the one presented here. See his Robinson (1994) and also Robinson (1988: 170-173).
97. This is reminiscent of Hylas's attempt to show that perceived objects cannot exist in the mind because they are perceived *at a distance*. (See Berkeley (1985: First Dialogue)). Compare also H.H. Price's reply to the charge that the world of the sense–datum theorist should be "flat": "Moreover, it is a strange misunderstanding to assume, as some eminent thinkers have, that visual sense-data must be "flat", i.e., two-dimensional. It is a plain phenomenological fact that visual fields have the property of depth. And why should the sense-datum philosophers, of all people, be supposed to deny this obvious fact?" (Price 1950: vii-viii). Contemporary projectivists about secondary qualities seem equally unfazed by the challenge of accounting for the apparent "*outness* or *distance*" suggested by these senses. See Baldwin (1992) and Boghossian and Velleman (1989).
98. For a discussion of the scientific feasibility of such goggles see Hans Moravec's discussion of "magic glasses." (Moravec 1988: Chapter 3)
99. Shades of David Stoves haunting dictum on the role of "in" in metaphysics! See footnote 25.
100. The title of this section is taken from Dretske's way of making this point. He speaks of the obvious fact that "the qualities in terms of which we distinguish thoughts, feelings, and experiences from one another, are nowhere to be found in the head." (Dretske 1995: 151)
101. Nagel made a similar observation in Nagel (1974: fn. 10).
102. In fairness to the HRT I should note one respect in which the HRT seems far superior to the WRT. Taken in combination with introspectionism the relocation$_h$ of qualia makes for a much more plausible theory of consciousness

than their relocation$_w$. For to ask the inner sense to monitor the outer qualia is surely to ask too much. I discuss this issue in Section 8.4.2.

103. In Section 8.4.2 I will argue for a stronger conclusion: the WRT seriously compromises the introspectionist account of consciousness. While I cannot prove that the two doctrines are incompatible, the evidence points in this direction.

104. This high-level description applies naturally to the internalist versions of introspectionism defended by Rosenthal and Pollock and the representationalism of Dretske and Tye. The positions of Armstrong and Lycan which combine an introspective component with externalism about qualia are harder to fit into this picture. Later on (see Section 8.4.2) I hope to persuade you that the difficulty of fitting Armstrong and Lycan into this picture points to a weakness in their position rather than to a weakness in my classificatory scheme.

105. I expect no objections to the claim that such accounts can differ according to the stance they take on representations as qualia-bearers. But perhaps it will be argued that no answer to the having-problem can get by without relying on representations as qualia revealers. It might be argued that it is obvious that an unrepresented quale—i.e., a quale that is not revealed—is a quale that is not had. In Chapter Ten I hope to show how this connection between representing qualia and having qualia can be broken.

106. The exclusion from consideration of accounts that do not involve representations as means by which qualia are revealed should not be taken to show that these accounts are uninteresting. Note that the account proposed in the concluding chapter is one that tries to explain the *having* of qualia without appeal to representational activity.

107. To capture Christopher Maloney's position more accurately the heading of this colum should read: "Qualia not revealed by higher order representation."

108. Rosenthal writes this passage as a comment on Cartesianism. But as his discussion on pages 10 to 11 of Rosenthal (fc: Section III) makes clear, he sees a very close tie between the Cartesian notion that all mental states are conscious and the thesis that consciousness is intrinsic. It therefore seems appropriate to use this passage as a comment on the consequences of the intrinsicalness of consciousness.

109. The problem has also been raised and relentlessly pursued by Fred Dretske. See especially Dretske (1993), Dretske 1995: Chapter 4).

110. Of course there is more to Rosenthal's story. In particular, he ends his Rosenthal, (fc.) by presenting two ingenious arguments in support of his specific version of the HO approach: the argument from introspection and the argument from reporting and expressing. These arguments deserve serious attention. But for our purposes it suffices to note that both arguments are strictly "philosophical" in the sense that is relevant here. That is, they proceed from the sort of principles and intuitions that one tends to discover in the armchair rather than on the field trip or in the laboratory.

111. Ever since the publication of his Searle (1980), Searle has been tireless in his efforts to expose the computationalist magic trick of pulling consciousness (and intentionality) out of a hat that contains only formal symbols in motion.

The amazement I express at Maloney's theory of qualitative consciousness (as described thus far) is precisely of the Searlian kind.

112. The move is tempting; but it has not gone uncontested. Lycan (1987: 58-59) has presented a powerful argument to the effect that mixed theories of this sort—functionalism + physicalism, or computationalism + physicalism—are nonstarters. He argues that adding in a bit of remedial physicalism at the points where functionalism/computationalism becomes uncomfortably constricting does not help to alleviate the pressure. For on closer inspection it turns out that the added dollop of physicalism dissolves into functionalism leaving the original functionalist/computationalist position exactly where it was at the outset. Thus the maneuver buys no extra explanatory power.

113. If we discovered that one of these theories were true we should be in precisely the sort of situation that Cummins envisioned as a possibility—see footnote 87. Douglas Adams has parodied this intellectual predicament in unforgettable manner. On hearing Deep Thought—the second greatest computer of the Universe of Space and Time—announce that the Answer to the Great Question of Life, the Universe, and Everything is 42, the audience is understandably stunned. One may suppose that the audience experienced the sort of unease that I have been trying to make you feel about representation-based theories of consciousness, and in most exquisite form at that. Adams (1989) develops this and other philosophically relevant points in Chapters 25-28.

114. In Section 9.3 I will attack the Armstrong/Lycan solution of the having-problem from yet another angle. There the focus will be on the question how representing a relocated *celluloid* quale can lead to an experience that appears to have a set of features distinct from, and icompatible with, the features that celluloid qualia allegedly have. This problem is quite general. It arises for all representation-based accounts of consciousness that work with relocated celluloid qualia. My reason for targeting this critique at the Armstrong/Lycan approach is its unusual lucidity. Critique is what you get in return for writing clearly.

115. It is a little misleading to list Rosenthal in this context because he explicitly rejects the inner sense interpretation of introspection.

116. In an exchange with Rosenthal Armstrong makes a tantalizing comment that suggests just this sort of picture. Responding to Rosenthal's claim—a claim that Rosenthal seems to have taken to state the obvious—that "introspection [cannot] grasp the perceived properties of physical objects" (Rosenthal 1984: 109), Armstrong writes: "I think that [this claim] must be wrong." (Armstrong 1984b: 235) This certainly suggests the view that introspection can "reach out" to the external objects and the physical paint qualia they bear. Though the setting in which this exchange occurs keeps Armstrong from elaborating his position in the direction that interests us here, it seems reasonable to interpret him as saying that introspection can, as it were, extrospect.

117. The other thing you get in return is an enormous problem when it comes to determining the place of a qualia in a materialistically conceived world. I address this version of the nature-problem in Section 10.6, where I try to present a metaphysical model broad enough to accommodate qualia while retaining those features of materialism that are worth wanting.

118. This is a little misleading for I will not focus exclusively on the intrinsic features of qualia. A big part of my account will depend on the relation of qualia to the experiencing subject (see Section 10.5.2). But it will do as a way to get the schematic idea across.
119. The problems arising from nonveridical experience are primarily relevant to the WRT. In Section 9.2.5 I indicated why and how these problems affect the HRT.
120. The term "sensation" is here used in a broad sense so as to cover the qualitative aspect of perceptions, hallucinations, dreams, etc.
121. Given Brentano's mind-body dualism, his intentional objects should, strictly speaking, be characterized as Mental. Hence they are as unfit to bear any kind of relocated qualia. The relocationist$_h$ would therefore be ill advised to pin her hope on intentional objects à la Brentano.
122. The issue whether any physical objects, actual or merely possible, can exemplify phenomenal properties is an issue that will be discussed in Section 9.3.
123. Unlike in the case of the argument from illusion no defense is needed, I hope, for the claim that the argument from atomism cuts equally against both versions of the relocation thesis. If there is a problem concerning the instantiation of qualia in material objects then it matters not which material objects are the favored qualia-bearers of one's theory. Neurons and tomatoes are equally ill suited to do the job.
124. See, for example, Sellars (1962), Aune (1967: 172-175), Delaney (1972), Hilbert (1987: 43-46), Lycan (1987: 93-111).
125. "A philosopher who regards ignorance of a scientific theory as a sufficient reason for not writing about it cannot be accused of complete lack of originality, as a study of recent philosophical literature will amply prove." (Broad 1969: 3) Foolishly ignoring Broad's advice, I will come back to the grain problem in Section 10.6.3.4, where I tentatively endorse a Russellian attempt to harmonize the simplicity and homogeneity of our qualia with the grainy structure of matter.
126. Or, perhaps, the enlargement of physicalism. This may be the point of Sellars's claim that the scientific account of matter will have to be enriched by the addition of Sensa. For a discussion of Sellarsian sensa see Lycan (1987: 93-111).
127. See Roderick Chisholm's (1957: Chapters 4 and 8) seminal discussion of the various uses of appear-words. Modified versions of this discussion appear in all of the editions of his important *Theory of Knowledge*—see, for example, Chisholm (1977: 26-30). Of particular interest is the distinction between the comparative and noncomparative use of appear-words. This distinction has been seized upon by numerous authors. See, for example, Pollock (1974: Chapter 4), Pollock (1986: 34-35), and Lehrer (1990: 67-68).
128. Questions concerning the extent of the "illusory element" in our experience of color are further explored in Section 9.3.6.
129. It is probably unfair to hold Armstrong responsible for the letter of the Gestaltists's metaphysical doctrines. But I believe that even the Gestaltist spirit is not much help for Armstrong's enterprise. I come back to this topic in Section 9.3.6.3.
130. This consideration is reminiscent of the phenomena discovered in the fascinating self-attribution experiments that the paper Nisbett and Wilson (1977)

introduced to the philosophical community. We might be taken to confabulate an acquaintance with simple colors so as to provide a rationalizing story for our otherwise incomprehensible ability to detect systematic relationships between the mere gaps or "that whiches" that colors are according to Armstrong.

131. I am inclined to think that this point, if it holds up, is of much broader significance. But this is not the place to explore the possible ramifications of the inconceivability of the world the DIRE realists would have us believe in.
132. Since, in this instance, we are talking about phenomena—the way things seem—saving the phenomena is saving all.
133. Announcing one's inability to understand *x* is a poor kind of criticism of *x*. True enough. But perhaps it suffices to shift the burden of proof: it is now incumbent upon the proponent of the obscure thesis to explain himself more clearly, if he can.
134. See Lycan (1987: 90-91) for a discussion of this issue. Interestingly he thinks that Armstrong's direct realism offers a solution of this problem. As I have tried to develop this problem within the framework of Armstrong's theory, I doubt that an appeal to Armstrong's direct realism can help matters here.
135. I choose the expression "self-presenting" to evoke two ideas. First, the expression serves as a reminder that having a quale is not a matter of the conscious subject's acting upon it in any way. There is the conscious subject, there is the quale, and there is nothing that the former *does* in order to have the latter. The second idea that the expression "self-presenting" helps to recall is that the quale *itself* (not some stand-in or representation of the quale) is present to us when we have it. Were one to insist on representation talk I could put the same point by saying that the quale is its own representation.
136. In developing my account of consciousness as the having of qualia I freely help myself to the bundle theory. I do this without so much as stopping to address most of the many objections that have been raised against this theory. Lack of space and, more importantly, the lack of philosophical expertise keep me from attempting to develop a full-featured and defensible version of the bundle theory at this point. The assumption that a viable version of the bundle theory can be worked out is, then, one of the major liabilities that my account of consciousness is burdened with.
137. There may be one way of resisting this conclusion. One might argue that the adverbialist construal of consciousness is an analysis of what it is to have qualia. The borderline between elimination and reduction is fuzzy. But in this case I feel sure that we are dealing with elimination, not reduction.
138. I am pleased to report that after this was written (though not, regrettably, because this was written) two of the leading adverbialists recanted. Chapter 13 of Chisholm's (1996) is a defense of the thesis that appearances are individual things where this is understood to be a view incompatible with adverbialism. Tye, who not so long ago wrote the most systematic defense of adverbialism to date (see Tye (1989)), now suspects that in adverbialism the dreaded phenomenal objects with their qualia might just have been "hidden under a verbal smoke screen." (Tye 1995a: 74)
139. This is, of course, the old problem plaguing the colored-brain thesis. See Section 7.5.

140. Some of the most vocal defenders of the neutral theory of experience interpreted this neutrality as the neutrality of experience between mind and matter. But neutralism is only a part of the much more comprehensive doctrine of neutral monism. Neutral monists like, for example, William James, Ernst Mach, and Bertrand Russell took their theory to be a comprehensive metaphysical system on a par with such metaphysical doctrines as materialism, idealism, or dualism. Neutral monism will reappear in the later sections of this chapter (see especially Section 10.6.4; see also Section 10.6.3.4). The position defended in this chapter has many points of contact with the metaphysics of neutral monism.
141. See also Russell (1919: 305-306) and especially "Consciousness and Experience" which is Chapter 12 of Russell (1975). In "Consciousness?" Russell (1979: Chapter 20) approaches this question in a quite different manner by undermining the act/object distinction from the side of the object. It is worth noting that before his conversion to the monistic view Russell himself provided a characteristically forceful statement of what may seem the best reason for upholding the act/object analysis. Reviewing William James's exposition of neutral monism, the as yet unconverted Russell wrote: "On grounds of the purest empiricism, from mere inspection of experience, I for my part should hold it obvious that perception is in its intrinsic nature a fact of relation, involving an act as well as an object." (Russell 1912: 574) And we have just seen that the same Russell later on felt sure, again on the grounds of purest empiricism, that the self is a fiction. Does this indicate that there is something wrong with Russell's philosophical perception? I think not. After having undergone a similarly dramatic *Gestalt* shift in his philosophical views, Chisholm made the following illuminating remark about his own conversion. He has his opponent say: "You say 'surely.' How, then could you have thought otherwise?" To which Chisholm replies: "The mistake, I fear, is one that is frequent in philosophy. For if an assumption is such that, if it is true, then it can be used to solve a great variety of difficult problems, then it is very difficult to contemplate the possibility that the assumption is false." (Chisholm 1979: 317) Of course, this rationale for radical change in view cuts both ways; the seductive appeal of the new view is also motivated, at least in part, by its problem-solving capacities.
142. This question may seem particularly pressing once it is realized that this is how Russell characterized the position of the American neutral monists and Mach. He classified them as realists "who only retain the object." (Russell 1978: 22)
143. Here is how Moritz Schlick puts the matter. Criticizing Locke's introspectionism writes (I substitute the term "percept" for Schlick's term "idea"): "Consciousness is not related to [percepts] as the stomach is related to the food that it takes in and digests. Indeed, it is [percepts] that constitute consciousness. They need not first be perceived by some special act; their very existence as data of consciousness is identical with their being perceived. For them, esse is the same as percipi. Hence there is no need to postulate a specific capacity to perceive the contents of consciousness..." (Schlick 1985: 131)
144. I borrow this use of the term "feeling" from Thomas Reid. Taken in this sense feeling "has no object; the feeling and the thing felt are one and the

same." (Reid 1785: 20) Reid is, of course, one of the classical proponents of monism, as is evidenced by the following quotation: "*Sensation* is a name given by philosophers to an act of mind which may be distinguished from all others by this, that it hath no object distinct from the act itself. Pain of every kind is an uneasy sensation. When I am pained, I cannot say that the pain I feel is one thing, and that my feeling it is another thing. They are one and the same thing, and cannot be disjoined, even in imagination. Pain, when it is not felt, has no existence...What we have said of pain may be applied to every other sensation." (Reid 1785: 18-19)

145. At this point a defender of the relational theory of experience might object as follows: The relational theory of experience is not wedded to the substance/attribute theory. Why should an adherent of the relational theory not appropriate the bundle theory of the sensory object and use it to solve the problems about qualia-bearers of which you made so much? This objection is correct, as far as it goes. You can have an act/object account of experience while offering a bundle theory of the experienced object. But this proposal addresses only one of the two main problems that I raised for relational theories. The other problem is to account for the having-relation. And as long as you stick to the act/object assumption, the having-problem persists undiminished by the adoption of the bundle theory.
146. The adoption of trope theory would be another plausible reaction to the problem of unborne qualities. A trope is an individual whose existence is, in principle, no more or less problematic than that of any other individual. I did not adopt trope theory because I wanted to take Russell as my guide in developing the neutral monist theory of consciousness; and Russell does not construe percepts as bundles of tropes. Russell's view is not easily classified. He is often portrayed as holding the view that individuals are bundles of universals. But he explicitly rejects this view when commenting on the writings from which I quote. Responding to a critic, he writes that he has "failed to understand the tentative theory, set forth in the *Inquiry*, according to which a given shade of colour is a particular, not a universal...My theory has been misunderstood because readers have persisted in regarding a given shade of colour as a universal. We are accustomed to the idea that a particular may persist through a finite continuous portion of space-time; what I maintain is that it may occupy a *dis*continuous portion of space time." (Russell 1946: 685) Later on Russell (1946: 714) makes it clear that he means to hold that any precisely defined quality is a particular. Though this view differs from the trope theory, Russell's particular qualities are like tropes in being particulars. And as long as we deal with particulars, questions concerning exemplification or instantiation do not arise.
147. We encountered the same phenomenon before when discussing the location of qualia and the location of the mind or of consciousness.
148. The translations of this and the following quotations from Planck are my own.
149. There are at least two interestingly different ways in which this entailment fails. First, there is the traditional objection that simply makes the logical point that even complete silence about colors on the part of the physicists, physiologists, etc., does not entail the nonexistence of colors on the external objects—see footnote 53. Second, and more relevantly, there is the position

that a DIRE realist about color might occupy. If colors are primary-property complexes, then there is no reason to believe that scientific explanations never mention colors. What is true is the uninteresting claim that scientific explanations never contain words from the traditional color vocabulary.

150. Russell sometimes made this point by saying "that space has six dimensions and not only three...what in the space of physics, counts as a point...is really a three-dimensional complex of which the total of one man's percepts is an instance." (Russell 1975: 79)
151. See, for example, Russell (1954: 320), Russell (1979: 111), Russell (1976: 229), Russell (1956: 151), Russell (1975: 19), Schlick (1985: 311-314), Feigl (1958: 473-474).
152. A more detailed discussion of these two versions of the identity theory is contained in my Stubenberg (1997). There I use the labels "the Australian version" and "the Austrian version" to mark the fact that the most prominent champions of the mainline version have been Australians (or serious Australophiles), while some of the more prominent defenders of the other identity theory have been Austrians. In conversation, David Chalmers has suggested to call the other identity theory "the reverse identity theory." As the other identity gains more definition it will become clear just how perceptive Chalmers's label is.
153. The identity theorist might also choose mental paint instead of physical paint; or she might give up on relocation altogether and "go adverbial." Though these are important distinctions within the mainline version of the identity theory, they do not tend to diminish the contrast between the mainline version of the identity theory and the other identity theory.
154. For an impressive list of different interpretations of Russell's neutral monism see the first paragraph of Tully (1988). The most systematic exposition and development of many of the pertinent topics of Russell's position is to be found in a series of papers by Grover Maxwell—Maxwell (1967), Maxwell (1970a), Maxwell (1970b), Maxwell (1972), Maxwell (1976).
155. Here I am assuming that the monist's causal inference to the existence of external objects is justified. The monistic literature is filled with attempts to defend this version of realism. The most extensive discussion is contained in Russell (1976). As a convert from phenomenalism, he seems particularly sensitive to the need and the difficulty of defending the realistic component of structural realism.
156. Discussing the place of Cartesian substances in the physical world Lockwood has produced the following arresting picture: "It remains obscure, on the dualist theory, just how the material is supposed to dovetail with immaterial mind. For, on the face of it, there are no mind-shaped gaps in the material fabric; the material world offers no explanatory or descriptive slots into which immaterial minds could comfortably fit. (One pictures matter saying to Cartesian mind: 'This universe ain't big enough for both of us'!)" (Lockwood 1993: 272) What I am suggesting is that while the gaps in matter may not be big enough to accommodate souls they may be just big enough to accommodate qualia. I should note that Lockwood is one of a handful of philosophers defending a view much like the one advocated here.
157. Russell expresses this basic idea in a number of places. See, for example, Russell (1976: 230-231), Russell (1954: 320), Russell (1957: 121).

158. This suggestion can be found in Maxwell (1976: 355) and in (Aune 1989: 10-14).
159. It is worth noting that the one point that gives him pause is the knowledge argument—an argument that, after having been rediscovered more than half a century after Russell first presented it, now finally receives the attention it deserves. For Russell's version of the argument see Russell (1954: Chapter 37).

References

Ackerman, Felicia. 1987. "An Argument for a Modified Russellian Principle of Acquaintance." *Philosophical Perspectives* I, Metaphysics. Pages: 501-512.

Adams, Douglas. 1989. *The Hitchhiker's Guide to the Galaxy*. First published in 1979. New York: Harmony Books.

Alexander, Samuel. 1914. "The Basis of Realism." *Proceedings of the British Academy* . Reprinted in Chisholm (1960). Pages: 186-222. Page references are to this reprinting.

Alston, William P. 1991. *Perceiving God. The Epistemology of Religious Experience*. Ithaca: Cornell University Press.

Anscombe, G.E.M. 1959. *An Introduction to Wittgenstein's Tractatus*. London: Hutchinson.

Armstrong, D.M. 1961. *Perception and the Physical World*. London: Routledge & Kegan Paul.

Armstrong, D.M. 1963. "Is Introspective Knowledge Incorrigible?" *Philosophical Review* 72. Pages: 417-432.

Armstrong, D.M. 1966. "The Nature of Mind." *Arts (Proceedings of the Sydney University Arts Association)* 3, 1. Pages: 37-48. Reprinted in Armstrong (1980). Pages: 1-15. Page references are to this reprinting.

Armstrong, D.M. 1968. *A Materialist Theory of Mind*. London: Routledge & Kegan Paul.

Armstrong, D.M. 1973. "Epistemological Foundations for a Materialist Theory of the Mind." *Philosophy of Science* 40. Pages: 178-93. Reprinted in Armstrong (1980). Pages: 32-54. Page references are to this reprinting.

Armstrong, D.M. 1977. "The Causal Theory of Mind." *Neue Hefte für Philosophie* 11. Pages: 82-95. Reprinted in Armstrong (1980). Pages: 16-31. Page references are to this reprinting.

Armstrong, D.M. 1978. "What Is Consciousness?" *Proceedings of the Russellian Society (Sydney)* 3. Pages: 65-76. Reprinted in Armstrong (1980). Pages: 55-67. Page references are to this reprinting.

Armstrong, D.M. 1979. "Perception, Sense-Data and Causality." Originally in Macdonald (1979). Reprinted in Armstrong (1980). Pages: 132-148. Page references are to this reprinting.

Armstrong, D.M. 1980. *The Nature of Mind*. Brighton, Sussex: Harvester Press.

Armstrong, D.M. 1984a. "Consciousness and Causality." In Armstrong and Malcolm (1984). Pages: 103-191.

Armstrong, D.M. 1984b. "Replies." In Bogdan (1984). Pages: 225-269.

Armstrong, D.M. 1984c. "Self-Profile." In Bogdan (1984). Pages: 3-51.

Armstrong, D.M. and Norman Malcolm. 1984. *Consciousness and Causality. A Debate on the Nature of Mind*. Oxford: Basil Blackwell.

Armstrong, D.M. 1987. "Smart and the Secondary Qualities." In Pettit, et al. (1987). Pages: 1-15.

Aune, Bruce. 1967. *Knowledge, Mind, and Nature*. Atascadero: Ridgeview Publishing Company.

Aune, Bruce. 1989. "Maxwell's Structural Realism." In Maxwell and Savage (1989). Pages: 1-15.

Ayer, A.J. 1971. *Russell and Moore. The Analytical Heritage*. London: Macmillan.

Ayer, A.J. 1986. *Freedom and Morality and Other Essays*. First published in 1984. Oxford: Oxford University Press.

Baldwin, James Mark, ed. 1901. *Dictionary of Philosophy and Psychology*. New York: The Macmillan Company.

Baldwin, Thomas. 1992. "The Projective Theory of Sensory Content." In Crane (1992). Pages: 177-195.

Barnes, Jonathan. 1982. *The Presocratic Philosophers*. London: Routledge & Kegan Paul.

Berkeley, George. 1985. *Three Dialogues Between Hylas and Philonous*. First published in 1713. Edited by R.M. Adams. Indianapolis: Hackett.

Block, Ned, ed. 1980. *Readings in Philosophy of Psychology*, Volume 1. Cambridge, MA: Harvard University Press.

Block, Ned, Owen Flanagan and Güven Güzeldre, eds. fc. *The Nature of Consciousness: Philosophical Debates*. Cambridge: MIT.

Block, Ned. 1978. "Troubles With Functionalism." Originally in Savage (1978). Reprinted in Block (1980), Vol. 1. Pages: 268-305. Page references are to this reprinting.

Block, Ned. 1992. "Begging the Question against Phenomenal Consciousness." *Behavioral and Brain Sciences* 15, 2. Pages: 205-206.

Block, Ned. 1994. "Consciousness." In Guttenplan (1994). Pages: 210-219.

Block, Ned. 1995. "On a Confusion about a Function of Consciousness." *Behavioral and Brain Sciences* 18, 2. Pages: 227-247.

Bogdan, Radu J., ed. 1984. *D.M. Armstrong*. Dordrecht: Reidel.

Boghossian, Paul A. and David J. Velleman. 1989. "Color as Secondary Quality." *Mind* 98, 389. Pages: 81-103.

Borst, C.V., ed. 1970. *The Mind-Brain Identity Theory*. London: The Macmillan Press.

Brentano, Franz Clemens. 1973. *Psychology From an Empirical Standpoint*. Edited by Oskar Kraus; translated from the German by Antos C. Rancurello, D. B. Terrell and Linda L. McAlister; English edition edited by Linda L. McAlister. London: Routledge and Kegan Paul.

Broad, C.D. 1919. "Is there "Knowledge by Acquaintance"?" *Proceedings of the Aristotelian Society* Supplementary Volume II. Part of the PAS Symposium: Is there "Knowledge by Acquaintance"? Pages: 206-220.

Broad, C.D. 1969. *Scientific Thought*. First published in 1923. New York: Humanities Press.

Brockman, John, ed. 1990. *Speculations*. New York, NY: Prentice Hall Press.

Butchvarov, Panayot. 1980. "Adverbial Theories of Consciousness." In French, et al. (1980), Volume 5. Pages: 261-280.

Campbell, Keith. 1984. *Body and Mind.* (2nd Edition). Notre Dame: University of Notre Dame Press.

Carruthers, Peter. 1989. "Brute Experience." *The Journal of Philosophy* 86. Pages: 258-269.
Carruthers, Peter. 1992. "Consciousness and Concepts." *The Aristotelian Society. Supplementary Volume* LXVI. Pages: 41-59.
Castañeda, Hector-Neri. 1979. "Philosophical Method and Direct Awareness of the Self." In Sosa (1979). Pages: 1-58.
Casati, Roberto and Graham White, eds. 1993. *Philosophy and the Cognitive Sciences.* Papers of the 16th International Wittgenstein Symposium. Kirchberg: The Austrian Ludwig Wittgenstein Society.
Chalmers, David. 1994. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies* 2. Also in. Pages: 200-219.
Chalmers, David. 1996. *The Conscious Mind. In Search of a Fundamental Theory.* New York: Oxford University Press.
Chisholm, Roderick M. 1957. *Perceiving. A Philosophical Study*. Ithaca: Cornell University Press.
Chisholm, Roderick M. 1966. *Theory of Knowledge*. (1st Edition). Englewood Cliffs, NJ: Prentice Hall.
Chisholm, Roderick M. 1970. "Brentano on Descriptive Psychology and the Intentional." In Morick (1970). Pages: 130-149.
Chisholm, Roderick M. 1973. "Homeless Objects." In Chisholm (1982). Pages: 37-52.
Chisholm, Roderick M. 1977. *Theory of Knowledge*. Second Edition. Englewood Cliffs: Prentice-Hall.
Chisholm, Roderick M. 1979. "Objects and Persons: Revisions and Replies." In Sosa (1979). Pages: 317-388.
Chisholm, Roderick M. 1982. "Beyond Being and Nonbeing." In Chisholm (1982a). Pages: 53-67.
Chisholm, Roderick M. 1982a. *Brentano and Meinong Studies*. Atlantic Highlands: Humanities Press.
Chisholm, Roderick M. 1983. "Believing as an Intentional Concept." In Parret (1983). Pages: 48-56.
Chisholm, Roderick M. 1989. *Theory of Knowledge*. Third Edition. Englewood Cliffs: Prentice Hall.
Chisholm, Roderick M. 1996. *A Realistic Theory of Categories. An Essay in Ontology*. Cambridge: Cambridge University Press.
Chisholm, Roderick M., ed. 1960. *Realism and the Background of Phenomenology*. Atascadero: Ridgeview Publishing Company.
Churchland, Patricia S. 1983. "Consciousness: The Transmutation of a Concept." *Pacific Philosophical Quarterly* 64. Pages: 80-95.
Churchland, Paul M. and Patricia S. Churchland. 1981. "Functionalism, Qualia, and Intentionality." *Philosophical Topics* 12. Pages: 121-145.
Churchland, Paul. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.
Colodny, R., ed. 1970. *Pittsburgh Studies in the Philosophy of Science*, Volume IV. Pittsburgh: University of Pittsburgh Press.
Colodny, Robert, ed. 1962. *Frontiers of Science and Philosophy*. Pittsburgh: Pittsburgh University Press.

Crane, Tim. 1992. *The Contents of Experience*. Cambridge: Cambridge University Press.

Crane, Tim. 1995. *The Mechanical Mind*. London: Penguin Books.

Crick, Francis. 1994. *The Astonishing Hypothesis. The Scientific Search for the Soul*. New York: Charles's Scribener's Sons.

Cummins, Robert. 1983. *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press. A Bradford Book.

Cummins, Robert. 1984. "The Mind of the Matter." *PSA* 2. Pages: 791-798.

Cummins, Robert. 1989. *Meaning and Mental Representation*. Cambridge, Massachusetts: MIT Press.

Davidson, R.J., G.E. Schwarz and D. Shapiro, eds. 1983. *Consciousness and Self-Regulation*, Volume 3. Plenum Publishing Corporation.

Davies, Martin and Glyn W. Humphreys, eds. 1993. *Consciousness. Psychological and Philosophical Essays*. Oxford: Blackwell.

Delaney, Cornelius F. 1972. "Sellars's Grain Argument." *Australasian Journal of Philosophy* 50, 1. Pages: 14-16.

Dennett, Daniel C. 1978. "Toward a Cognitive Theory of Consciousness." Originally in Savage (1978). Reprinted in Dennett (1981). Pages: 149-173. Page references are to this reprinting.

Dennett, Daniel C. 1979. "On the Absence of Phenomenology." In Gustavson and Tapscott (1979). Pages: 93-113.

Dennett, Daniel C. 1981. "Where Am I?" In Dennett (1981a). Pages: 310-323.

Dennett, Daniel C. 1981a. *Brainstorms. Philosophical Essays on Mind and Psychology*. Brighton, Sussex: Harvester Press.

Dennett, Daniel C. 1982. "How to Study Human Consciousness Empirically or Nothing Comes to Mind." *Synthese* 53. Pages: 159-180.

Dennett, Daniel C. 1987. *The Intentional Stance*. Cambridge, Mass.: MIT Press.

Dennett, Daniel C. 1987a. "Consciousness." In Gregory (1987). Pages: 160-164.

Dennett, Daniel C. 1987b. "Setting off on the Right Foot." In Dennett (1987). Pages: 1-11.

Dennett, Daniel C. 1987c. "Three Kinds of Intentional Psychology." In Dennett (1987). First published in 1981. Pages: 43-68.

Dennett, Daniel C. 1988. "Quining Qualia." In Marcel and Bisiach (1988). Pages: 42-77.

Dennett, Daniel C. 1990. "The Evolution of Consciousness." In Brockman (1990). Pages: 85-108.

Dennett, Daniel C. 1991. *Consciousness Explained*. Boston: Little, Brown and Company.

Dennett, Daniel C. 1993. "Living on the Edge." *Inquiry* 36, 1/2. Replies to papers contributed to a symposium on Dennett (1991). Pages: 135-159.

Dretske, Fred. 1969. *Seeing and Knowing*. Chicago: The University of Chicago Press.

Dretske, Fred. 1993. "Conscious Experience." *Mind* 102. Pages: 263-283.

Dretske, Fred. 1995. *Naturalizing the Mind*. Cambridge: MIT Press.

Eddington, Arthur Stanley. 1929. *The Nature of the Physical World*. New York: Macmillan.

Edgell, Beatrice. 1919. "Is there "Knowledge by Acquaintance"?" *Proceedings of the Aristotelian Society* Supplementary Volume II. Part of the PAS Symposium: Is there "Knowledge by Acquaintance" Pages: 194-205.

Farrell, B.A. 1950. "Experience." *Mind* 59. Pages: 170-198.

Feigl, Herbert, Michael Scriven and Grover Maxwell, eds. 1958. *Concepts, Theories, and the Mind-Body Problem. Minnesota Studies in the Philosophy of Science*, Volume II. Minneapolis: University of Minnesota Press.

Feigl, Herbert. 1958. "The "Mental" and the "Physical"." In Feigl, et al. (1958). Pages: 370-497.

Feigl, Herbert. 1967. "Postscript After Ten Years." In Feigl (1967a). Pages: 135-169.

Feigl, Herbert. 1967a. *The Mental and the Physical. The Essay and the Postscript.* Minneapolis: University of Minnesota Press.

Feigl, Herbert. 1970. "Mind-Body, *Not* a Pseudo-Problem." Originally in Hook (1966). Reprinted in Borst (1970). Pages: 33-42. Page references are to this reprinting.

Flanagan, Owen. 1991. *The Science of Mind*. Second Edition. Cambridge, Massachusetts: MIT Press.

Flanagan, Owen. 1992. *Consciousness Reconsidered*. Cambridge: MIT Press.

Fodor, Jerry. 1987. *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*. Cambridge: MIT Press.

Fodor, Jerry. 1989. "Making Mind Matter More." *Philosophical Topics* 17, 1. Pages: 59-79.

Fodor, Jerry. 1992. "The Big Idea." *The Times Literary Supplement* , July 3. Pages: 5-8.

Fodor, Jerry. 1994. *The Elm and the Expert. Mentalese and Its Semantics*. Cambridge: MIT Press.

French, Peter A., Theodore E. Jr. Uehling and Howard K. Wettstein, eds. 1980. *Studies in Epistemology. Midwest Studies in Philosophy*, Volume 5. Minneapolis: University of Minnesota Press.

French, Peter A., Theodore E. Jr. Uehling and Howard K. Wettstein, eds. 1986. *Studies in the Philosophy of Mind. Midwest Studies in Philosophy*, Volume 10. Minneapolis: University of Minnesota Press.

Geach, Peter. 1969. "Immortality." In Geach (1969a). Pages: 17-29.

Geach, Peter. 1969a. *God and the Soul*. London: Routledge and Kegan Paul.

Globus, G., G. Maxwell and I. Savodnik. 1976. *Consciousness and the Brain*. New York: Plenum Press.

Goldman, Alvin. 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, Alvin. 1993a. "Consciousness, Folk Psychology, and Cognitive Science." *Consciousness and Cognition* 2. Pages: 364-382.

Goldman, Alvin. 1993b. "The Psychology of Folk Psychology." *Behavioral and Brain Sciences* 16, 1. With Open Peer Commentary. Pages: 15-28.

Graham, George and Lynn G. Stephens. 1985. "Are Qualia a Pain in the Neck for Functionalists?" *American Philosophical Quarterly* 22, 1. Pages: 73-80.

Gregory, Richard, ed. 1987. *The Oxford Companion to the Mind*. Oxford: Oxford University Press.

Griffin, Donald R. 1984. *Animal Thinking*. Cambridge: Harvard University Press.
Gunderson, Keith, ed. 1975. *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.
Gunderson, Keith. 1985. *Mentality and Machines*. Second Edition. Minneapolis: University of Minnesota Press.
Gustavson, D.F. and B.L. Tapscott, eds. 1979. *Body, Mind, and Method*. Dordrecht: Reidel.
Guttenplan, Samuel, ed. 1994. *A Companion to the Philosophy of Mind*. Blackwell Companions to Philosophy. Oxford: Blackwell.
Güzeldere, Güven. 1995. "Is Consciousness the Perception of What Passes in One's Own Mind?" In Metzinger (1995). Pages: 335-358.
Hameroff, Stuart.R., Alfred.W. Kaszniak and Alwyn.C. Scott, eds. 1996. *Toward a Scientific Basis of Consciousness: The First Tucson Discussion and Debates*. Cambridge: MIT Press.
Hannay, Alastair. 1990. *Human Consciousness*. London: Routledge.
Hardin, C.L. 1988. *Color for Philosophers. Unweaving the Rainbow*. Indianapolis: Hackett.
Harman, Gilbert. 1989. "Some Philosophical Issues in Cognitive Science: Qualia, Intentionality, and the Mind-Body Problem." In Posner (1989). Pages: 831-848.
Harman, Gilbert. 1990. "The Intrinsic Quality of Experience." In Tomberlin (1990), Volume 4. Pages: 31-52.
Haugeland, John. 1981. "Semantic Engines." In Haugeland (1981a). Pages: 1-34.
Haugeland, John, ed. 1981a. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA: MIT Press. A Bradford Book.
Hess, Peter H. 1988. *Thought and Experience*. Toronto: University of Toronto Press.
Hicks, Dawes G. 1919. "Is there "Knowledge by Acquaintance"?" *Proceedings of the Aristotelian Society* Supplementary Volume II. Part of the PAS Symposium: Is there "Knowledge by Acquaintance"? Pages: 159-178.
Hilbert, David R. 1987. *Color and Color Perception. A Study in Anthropocentric Realism*. Stanford: CSLI Publications.
Hill, Christopher S. 1991. *Sensations. A Defence of Type Materialism*. Cambridge: Cambridge University Press.
Hofstadter, Douglas R. and Daniel C. Dennett, eds. 1981. *The Mind's I. Fantasies and Reflections on Self and Soul*. New York: Basic Books.
Holt, Edwin B. 1914. *The Concept of Consciousness*. London: George Allen & Company.
Hook, Sidney, ed. 1966. *Dimensions of Mind. A Symposium*. First published in 1960. New York: Collier Books.
Hume, David. 1973. *A Treatise of Human Nature*. First published in 1739. London: Oxford University Press.
Humphrey, Nicholas and L. Weiskrantz. 1967. "Vision in Monkeys after Removal of the Striate Cortex." *Nature* 215, August 5. Pages: 595-597.
Humphrey, Nicholas. 1972. "Seeing and Nothingness." *New Scientist* 30, March. Pages: 682-684.

Humphrey, Nicholas. 1974. "Vision in a Monkey without Striate Cortex: a Case Study." *Perception* 3. Pages: 241-255.
Humphrey, Nicholas. 1980. "Nature's Psychologists." Originally in Josephston and Ramachandran (1980). Reprinted in Humphrey (1984). Pages: 29-41. Page references are to this reprinting.
Humphrey, Nicholas. 1984. *Consciousness Regained. Chapters in the Development of Mind*. Oxford: Oxford University Press.
Humphrey, Nicholas. 1986. *The Inner Eye*. London: Faber and Faber.
Humphrey, Nicholas. 1992. *A History of the Mind*. New York: Simon & Schuster.
Jaynes, Julian. 1976. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Boston: Houghton and Mifflin.
Josephston, Brian and V. Ramachandran, eds. 1980. *Consciousness and the Physical World*. Pergamon Press.
Kaila, Eino. 1942. "Physikalismus und Phänomenalismus." *Theoria* 8. Pages: 85-125.
Kihlstrom, John F. 1987. "The Cognitive Unconscious." *Science* 237, 18 September. Pages: 1445-1452.
Kim, Jaegwon. 1996. *Philosophy of Mind*. Boulder: Westview Press.
Kneale, William. 1969. "Review of Armstrong's *A Materialist Theory of Mind*." *Mind* 78. Pages: 292-301.
Lakatos, I. and A. Musgrave, eds. 1967. *Problems in the Philosophy of Science*. Amsterdam: North-Holland.
Landesman, Charles. 1989. *Color and Consciousness. An Essay in Metaphysics*. Philadelphia: Temple University Press.
Lehrer, Keith and Johann Christian Marek, eds. 1997. *Austrian Philosophy, Past and Present: Essays in Honour of Rudolf Haller*. Boston Studies in the Philosophy of Science, Volume 190. Dordrecht: Kluwer.
Lehrer, Keith. 1990. *Theory of Knowledge*. Boulder and San Francisco: Westview Press.
Lewis, David. 1995. "Should a Materialist Believe in Qualia?" *Australasian Journal of Philosophy* 73, 1. Pages: 140-144.
Lloyd, Dan. 1989. *Simple Minds*. Cambridge, MA: MIT Press. A Bradford Book.
Locke, John. 1982. *An Essay Concerning Human Understanding*. First published in 1689. Oxford: Oxford University Press.
Lockwood, Michael. 1989. *Mind, Brain and the Quantum. The Compound 'I'*. Oxford: Basil Blackwell.
Lockwood, Michael. 1993. "The Grain Problem." In Robinson (1993). Pages: 271-291.
Loux, Michael J., ed. 1979. *The Possible and the Actual. Readings in the Metaphysics of Modality*. Ithaca: Cornell University Press.
Lycan, William G. 1987. *Consciousness*. Cambridge, MA: MIT Press. A Bradford Book.
Lycan, William G. 1990. "What Is the "Subjectivity of the Mental?"." In Tomberlin (1990), Volume 4. Pages: 109-130.
Lycan, William G. 1992. "True Colors." 24 pages. Unpublished Manuscript.
Lycan, William G. 1995a. "A Limited Defence of Phenomenal Information." In Metzinger (1995). Pages: 243-258.

Lycan, William G. 1995b. "Consciousness as Internal Monitoring, I." In Tomberlin (1995). Pages: 1-14.
Lycan, William G. 1995c. "We've Only Just Begun." *Behavioral and Brain Sciences* 18, 2. Comment on Block (1995). Pages: 262-263.
Lycan, William G. 1996. *Consciousness and Experience*. Cambridge: MIT Press.
Lycan, William G. fc. "Consciousness as Internal Monitoring." In Block, et al., (fc.).
Lyons, William. 1986. *The Disappearance of Introspection*. Cambridge, MA: MIT Press. A Bradford Book.
Macdonald, G.F., ed. 1979. *Perception and Identity. Essays Presented to A.J. Ayer with His Replies*. Ithaca: Cornell University Press.
Mackie, J.L. 1976. *Problems from Locke*. Oxford: Clarendon Press.
Mackie, J.L. 1977. "Problems of Intentionality." Originally in Pivcevic (1977). Reprinted in Mackie (1985). Pages: 102-116. Page references are to this reprinting.
Mackie, J.L. 1985. *Logic and Knowledge. Selected Papers*. Volume 1. Oxford: Clarendon Press.
Maloney, J. Christopher. 1985. "About Being a Bat." *Australasian Journal of Philosophy* 63, 1. Pages: 26-49.
Marcel, A. and E. Bisiach, eds. 1988. *Consciousness in Contemporary Science*. Oxford: Oxford University Press.
Maudlin, Tim. 1989. "Computation and Consciousness." *The Journal of Philosophy* 86. Pages: 407-432.
Maxwell, Glover. 1976. "Scientific Results and the Mind-Body Issue: Some Afterthoughts." In Globus, et al. (1976). Pages: 329-358.
Maxwell, Grover. 1967. "Scientific Methodology and the Causal Theory of Perception." In Lakatos and Musgrave (1967). With commentaries by W.V. Quine, K. R. Popper, A.J. Ayer, W.C. Kneale and Maxwell's replies. Pages: 148-160.
Maxwell, Grover. 1970a. "Structural Realism and the Meaning of Theoretical Terms." In Radner and Winokur (1970). Pages: 181-192.
Maxwell, Grover. 1970b. "Theories, Perception and Structural Realism." In Colodny (1970). Pages: 3-34.
Maxwell, Grover. 1972. "Russell on Perception: A Study in Philosophical Method." In Pears (1972). Pages: 110-146.
Maxwell, Mary Lou and Wade Savage, eds. 1989. *Science, Mind, and Psychology. Essays in Honour of Grover Maxwell*. Lanham: University Press of America.
McGinn, Colin. 1982. *The Character of Mind*. Oxford: Oxford University Press.
McGinn, Colin. 1988. "Consciousness and Content." *Proceedings of the British Academy*, 74. Pages: 219-39. Reprinted in McGinn (1991). Pages: 23-43. Page references are to this reprinting.
McGinn, Colin. 1989. "Can We Solve the Mind-Body Problem?" *Mind* 98. Pages: 349-366.
McGinn, Colin. 1991. "Consciousness and the Natural Order." In McGinn (1991a). Chapter 4 of this book. Pages: 44-88.
McGinn, Colin. 1991a. *The Problem of Consciousness. Essays Towards a Resolution*. Oxford: Basil Blackwell.

Meinong, Alexius. 1907. *Über die Stellung der Gegenstandstheorie im System der Wissenschaften*. Leipzig: R. Voigtländer.

Metzinger, Thomas, ed. 1995. *Conscious Experience*. Thorverton: Imprint Academic.

Moore, G.E. 1903. "The Refutation of Idealism." *Mind* 12. Pages: 433-453. Reprinted in Moore, G.E. 1951. Philosophical Studies. First published in 1922. New York: The Humanities Press. Pages: 1-30. Page references are to this reprinting.

Moore, G.E. 1919. "Is there "Knowledge by Acquaintance"?" *Proceedings of the Aristotelian Society* Supplementary Volume II. Part of the PAS Symposium: Is there "Knowledge by Acquaintance"? Pages: 179-193.

Moore, G.E. 1951. *Philosophical Studies*. First published in 1921. New York: The Humanities Press.

Moore, G.E. 1966. *Some Main Problems of Philosophy*. First published in 1953. New York: Collier Books.

Moravec, Hans. 1988. *Mind Children. The Future of Robot and Human Intelligence*. Cambridge: Harvard University Press.

Morick, Harold, ed. 1970. *Introduction to the Philosophy of Mind*. Glenview: Scott & Foresman.

Moser, Paul K. 1989. *Knowledge and Evidence*. Cambridge: Cambridge University Press.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *The Philosophical Review* 83. Pages: 435-450. Reprinted in Nagel (1979). Pages: 165-180. Page references are to this reprinting.

Nagel, Thomas. 1979. *Mortal Questions*. Cambridge, MA: Cambridge University Press.

Nagel, Thomas. 1979a. "Panpsychism." In Nagel (1979). Pages: 181-195.

Nagel, Thomas. 1979b. "Subjective and Objective." In Nagel (1979). Pages: 196-213.

Nagel, Thomas. 1986. *The View From Nowhere*. Oxford: Oxford University Press.

Nagel, Thomas. 1987. *What Does It All Mean? A Very Short Introduction to Philosophy*. New York, Oxford: Oxford University Press.

Natsoulas, Thomas. 1982. "Conscious Perception and the Paradox of "Blind Sight"." In Underwood (1982), Volume 3. Pages: 79-109.

Nisbett, R.E. and T.D. Wilson. 1977. "Telling More than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84. Pages: 231-59.

Parret, H., ed. 1983. *On Believing*. Berlin: DeGruyter.

Parsons, Terence. 1987. "Entities without Identity." *Philosophical Perspectives* 1. Pages: 1-19.

Pears, D.F., ed. 1972. *Bertrand Russell. A Collection of Critical Essays*. New York: Anchor Books.

Pettit, Phillip, Richard Sylvan and Jean Norman, eds. 1987. *Metaphysics and Morality: Essays in Honour of J.J.C. Smart*. Oxford: Basil Blackwell.

Pivcevic, Edo, ed. 1977. *Phenomenology and Philosophical Understanding*. Cambridge: Cambridge University Press.

Place, U.T. 1956. "Is Consciousness a Brain Process?" *The British Journal of Psychology* XLVII. Reprinted in Borst (1970). Pages: 42-51. Page references are to this reprinting.

Planck, Max. 1958. *Physikalische Abhandlungen und Vorträge*. Volume 3 Bände. Aus Anlaß seines 100. Geburtstages (23. April 1958) herausgegeben von dem Verband Deutscher Physikalischer Gesellschaften und der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. Braunschweig: Frederick Vieweg & Sohn.

Planck, Max. 1958a. "Die Einheit des Physikalischen Weltbildes." In Planck (1958), Volume 3. Vortrag, gehalten am 9. Dezember 1908 in der naturwissenschaftlichen Fakultät des Studentenkorps and der Universität Leiden. Pages: 6-29.

Planck, Max. 1958b. "Zwanzig Jahre Arbeit am Physikalischen Weltbild." In Planck (1958), Volume 3. Vortrag, gehalten am 18 Februar 1929 im Physikalischen Institut der Universität Leiden. Pages: 179-208.

Planck, Max. 1990. *Vom Wesen der Willensfreiheit*. Mit einer Einleitung von Armin Hermann. Frankfrut am Main: Fischer.

Planck, Max. 1990a. "Kausalgesetz und Willensfreiheit." In Planck (1990). Vortrag, gehalten in der Preußischen Akademie der Wissenschaften am 17. Februar 1923. Pages: 81-117.

Planck, Max. 1990b. "Sinn und Grenzen der exakten Wissenschaft." In Planck (1990). Vortrag, gehalten in Großen Hörsaal des Physikalischen Instituts der Universität Bonn am 28. März 1947. Pages: 228-249.

Plantinga, Alvin. 1976. "Actualism and Possible Worlds." *Theoria* 42. Pages: 139-160. Reprinted in Loux (1979). Pages: 253-273. Page references are to this reprinting.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford: Oxford University Press.

Pollock, John L. 1989. *How to Build a Person. A Prolegomenon*. Cambridge, MA: MIT Press. A Bradford Book.

Pollock, John. 1974. *Knowledge and Justification*. Princeton: Princeton University Press.

Pollock, John. 1986. *Contemporary Theories of Knowledge*. Totowa: Rowman & Littlefield.

Posner, Michael I., ed. 1989. *Foundations of Cognitive Science*. Cambridge, Mass: MIT Press.

Price, H.H. 1950. *Perception*. First published 1932. London: Methuen & Co.

Putnam, Hilary. 1975. "The Meaning of "Meaning"." Originally in Gunderson (1975). Reprinted in Putnam (1975). Pages: 215-271. Page references are to this reprinting.

Putnam, Hilary. 1975. *Mind, Language and Reality*. Philosophical Papers Volume 2. Cambridge: Cambridge University Press.

Radner, Daisie and Michael Radner. 1989. *Animal Consciousness*. Buffalo: Prometheus Books.

Radner, M. and S. Winokur, eds. 1970. *Minnesota Studies in the Philosophy of Science*, Volume IV. Minneapolis: University of Minnesota Press.

Reid, Thomas. 1785. *Essays on the Intellectual Powers of Man*. Edited and abridged by A.D. Woozley. (No publication date). Charlottesville: Ibis Publishing.

Rey, Georges. 1983. "A Reason for Doubting the Existence of Consciousness." In Davidson, et al. (1983), Volume 3. Pages: 1-39.

Robinson, Howard, ed. 1993. *Objections to Physicalism*. Oxford: Oxford University Press.

Robinson, Howard. 1993. "Introduction." In Robinson (1993). Pages: 1-25.

Robinson, William S. 1988. *Brains and People. An Essay on Mentality and Its Causal Conditions*. Philadelphia: Temple University Press.

Robinson, William S. 1994. "Intrinsic Consciousness." 23 pages. Unpublished Manuscript.

Rorty, Richard. 1965. "Mind-Body Identity, Privacy, and Categories." *The Review of Metaphysics* 19. Pages: 24-54. Reprinted in Borst (1970). Pages: 187-213. Page references are to this reprinting.

Rosenthal, David M. 1984. "Armstrong's Causal Theory of the Mind." In Bogdan (1984). Pages: 79-120.

Rosenthal, David M. 1986. "Two Concepts of Consciousness." *Philosophical Studies* 49. Pages: 329-359.

Rosenthal, David M. 1991. "The Independence of Consciousness and Sensory Quality." In Villanueva (1991). Pages: 15-36.

Rosenthal, David M. 1993. "Thinking that One Thinks." In Davies and Humphreys (1993). Pages: 197-223.

Rosenthal, David M. fc. "A Theory of Consciousness." In Block, et al., (fc.). An earlier version of this paper was published in 1990 as Report no. 40, Research Group on Mind and Brain, ZiF, University of Bielefeld. Pages:

Russell, Bertrand. 1912. "Review of William James *Essays in Radical Empiricism*." *Mind* 21. Pages: 571-575.

Russell, Bertrand. 1914. "On the Nature of Acquaintance." *Monist* 24. Pages: 1-16, 161-187, 435-453. Reprinted in Russell (1956b). Pages: 125-175. Page references are to this reprinting.

Russell, Bertrand. 1915. "Letter from Bertrand Russell." *The Journal of Philosophy, Psychology, and Scientific Methods* 12. Pages: 391-392.

Russell, Bertrand. 1918/19. "The Philosophy of Logical Atomism." *The Monist* 28/29. Reprinted in Russell (1977). Pages: 177-281. Page references are to this reprinting.

Russell, Bertrand. 1919. "On Propositions: What They Are and How They Mean." *Proceedings of the Aristotelian Society* 2. Pages: 1-43. Reprinted in Russell, (1977). Pages: 285-320. Page references are to this reprinting.

Russell, Bertrand. 1923. *The ABC of Atoms*. New York: Dutton & Company.

Russell, Bertrand. 1946. "Reply to My Critics." In Schilpp (1946). Pages: 679-741.

Russell, Bertrand. 1954. *The Analysis of Matter*. First published in 1927. New York: Dover.

Russell, Bertrand. 1956a. "Mind and Matter." In Russell (1956b). Pages: 135-153.

Russell, Bertrand. 1956b. *Portraits from Memory*. London: George Allen & Unwin.

Russell, Bertrand. 1957a. "Mentalism vs. Materialism." In Russell (1957b). Pages: 105-122.

Russell, Bertrand. 1957b. *Understanding History and Other Essays*. New York: Philosophical Library.

Russell, Bertrand. 1975. *My Philosophical Development*. First published in 1959. London: George Allen & Unwin.

Russell, Bertrand. 1976. *Human Knowledge. Its Scope and Limits*. First published in 1948. New York: Simon and Schuster.

Russell, Bertrand. 1977. *Logic and Knowledge. Essays 1901-1955*. First published in 1956. Edited by R.C. Marsh. London: George Allen & Unwin.

Russell, Bertrand. 1978. *The Analysis of Mind*. First published 1921. London: George Allen & Unwin.

Russell, Bertrand. 1979. *An Outline of Philosophy*. First published in 1927. London: Allen and Unwin.

Russell, Bertrand. 1980a. *An Inquiry into Meaning and Truth*. First published in 1940. London: George Allen & Unwin.

Russell, Bertrand. 1980b. *Our Knowledge of the External World*. Reprinting of the second edition of 1926. First edition 1914. London: George Allen & Unwin.

Russell, Bertrand. 1982. *The Problems of Philosophy*. First published in 1912. Oxford: Oxford University Press.

Russell, Bertrand. 1984. *Theory of Knowledge. The 1913 Manuscript*. Edited by Elizabeth Ramsden Eames in collaboration with Kenneth Blackwell. The Collected Papers of Bertrand Russell, Vol. 7. London: George Allen & Unwin.

Sanford, David H. 1984. "Armstrong's Theory of Perception." In Bogdan (1984). Pages: 55-78.

Savage, Wade C., ed. 1978. *Perception and Cognition. Issues in the Foundation of Psychology. Minnesota Studies in the Philosophy of Science*, Volume 9. Minneapolis: Minnesota Press.

Schilpp, Paul Arthur, ed. 1946. *The Philosophy of Bertrand Russell*, Volume V. The Library of Living Philosophers. Evanston, Ill: The Library of Living Philosophers, Inc.

Schilpp, Paul Arthur, ed. 1968. *The Philosophy of G.E. Moore*. La Salle, Illinois: The Open Court Publishing Co.

Schlick, Moritz. 1985. *General Theory of Knowledge*. First published in 1918. This is a translation of the revised edition of 1925. LaSalle: Open Court.

Searle, John R. 1980. "Minds, Brains, and Programs." *The Behavioral and Brain Sciences* 3. Pages: 417-424.

Searle, John R. 1980a. "Author's Response." *The Behavioral and Brain Sciences* 3. Pages: 450-456.

Searle, John R. 1987. "Indeterminacy, Empiricism, and the First Person." *The Journal of Philosophy* 84, 3. Pages: 123-146.

Searle, John R. 1989. "Consciousness, Unconsciousness, and Intentionality." *Philosophical Topics* 17. Pages: 193-209.

Searle, John R. 1992. *The Rediscovery of the Mind*. Cambridge: MIT Press.

Sellars, Wilfrid. 1962. "Philosophy and the Scientific Image of Man." Originally in Colodny (1962). Reprinted in Sellars (1963). Pages: 1-40. Page references are to this reprinting.

Sellars, Wilfrid. 1963. *Science, Perception and Reality*. London: Routledge & Kegan Paul.
Shoemaker, Sidney. 1975. "Functionalism and Qualia." *Philosophical Studies* 27. Pages: 291-315.
Shoemaker, Sydney. 1986. "Introspection and the Self." In French, et al. (1986), Volume 10. Pages: 101-120.
Smart, J.J.C. 1959. "Sensations and Brain Processes." *The Philosophical Review* 68. Reprinted in Borst (1970). Pages: 52-66. Page references are to this reprinting.
Smart, J.J.C. 1963. "Materialism." *The Journal of Philosophy* 60, 22. Pages: 651-662.
Sosa, Ernest, ed. 1979. *Essays on the Philosophy of Roderick M. Chisholm*. Amsterdam: Rodopi.
Sprigge, T.L.S. 1994. "Consciousness." *Synthese* 98. Pages: 73-93.
Sternberg, Robert J. 1995. *In Search of the Human Mind*. Orlando: Harcourt Brace & Company.
Stoppard, Tom. 1972. *Jumpers*. New York: Grove Press.
Stove, David. 1991. *The Plato Cult and other Philosophical Follies*. Oxford: Blackwell.
Stubenberg, Leopold. 1992a. "Alston's *Perceiving God*." 59 pages. Unpublished Manuscript.
Stubenberg, Leopold. 1992b. "What Is It Like to Be Oscar?" *Synthese* 90. Pages: 1-26.
Stubenberg, Leopold. 1993. "How to Study Consciousness: The First-Person vs. the Third-Person Perspective." In Casati (1993). Pages: 507-511.
Stubenberg, Leopold. 1996. "The Place of Qualia in the World of Science." In Hameroff, et al. (1996). Pages: 41-49.
Stubenberg, Leopold. 1997. "Austria vs. Australia: Two Versions of the Identity Theory." In Lehrer and Marek (1997). Pages: 125-146.
Tomberlin, James E., ed. 1990. *Action Theory and the Philosophy of Mind. Philosophical Perspectives*, Volume 4. Atascadero: Ridgeview.
Tomberlin, James E., ed. 1995. *AI, Connectionism and Philosophical Psychology. Philosophical Perspectives*, Volume 9. Atascadero: Ridgeview.
Tully, Robert. 1988. "Russell's Neutral Monism." *Russell* 8, 1-2. Pages: 209-224.
Twardowski, Kasimir. 1977. *On the Content and the Object of Presentations*. First published in 1894. Translated with and Introduction by Reinhardt Grossman. The Hague: Martinus Nijhoff.
Tye, Michael. 1989. *The Metaphysics of Mind*. Cambridge: Cambridge University Press.
Tye, Michael. 1993. "Reflections on Dennett and Consciousness." *Philosophy and Phenomenological Research* 53, 4. Pages: 893-898.
Tye, Michael. 1995a. *Ten Problems of Consciousness*. Cambridge: MIT Press.
Tye, Michael. 1995b. "What What It's Like is Really Like." *Analysis* 55. Pages: 125-126.
Underwood, Geoffrey, ed. 1982. *Aspects of Consciousness*, Volume 3. London:
Villanueva, Enrique, ed. 1991. *Consciousness*. Philosophical Issues, 1 1991. Atascadero: Ridgeview Publishing Company.

Weiskrantz, L. 1980. "Varieties of Residual Experience." *Quarterly Journal of Experimental Psychology* 32. Pages: 365-386.

Weiskrantz, Laurence. 1986. *Blindsight. A Case Study*. Oxford: Oxford University Press.

Whitehead, Alfred North. 1926. *The Concept of Nature*. First published in 1920. Cambridge: Cambridge University Press.

Wisdom, John. 1963. *Problems of Mind and Matter*. First published 1934. Cambridge: Cambridge University Press.

Wittgenstein, Ludwig. 1974. *Philosophical Investigations*. Oxford: Basil Blackwell.

Zihl, Josef. 1980. ""Blindsight": Improvement of Visually guided Eye Movements by Systematic Practice in Patients with Cerebral Blindness." *Neuropsychologia* 18. Pages: 71-77.

Detailed Table of Contents

1. Introduction 1
1.1 Philosophy or Science 1
1.2 Outline One: Some Key Ideas 2
1.3 Outline Two: Chapter Summaries 7
1.4 Caveats 10

2. Making our Ideas Clear: Consciousness and Qualia 12
2.1 The Thesis 12
2.2 Clarifying the Thesis 12
2.2.1 Qualitative Consciousness 13
2.2.2 An Apology for Qualia 18
2.2.2.1 The Datum 18
2.2.2.2 Why the Datum Is Obvious 19
2.2.2.3 The Neutrality Thesis 19
2.2.2.4 Qualia and Sense-Data 21
2.2.2.5 The Qualia Quiners 21
2.2.2.6 Why the Datum Is Important 24
2.2.2.7 Why the Datum Is Baffling 26
2.2.2.8 Deflating the Bafflement 27
2.2.2.9 Background Materialism 31
2.3 Two Problems: The Nature-problem and the Having-problem 33
2.4 Summary 35

3. On Method: The First-Person Perspective 36
3.1 The Principle of Phenomenological Adequacy 36
3.2 The Transformation of Consciousness Studies 38
3.3 The Attack on the First-Person Perspective 40
3.3.1 The Charge of Obscurantism and the Love of Mystery 40
3.3.1.1 The First-Person Perspective and anti-Materialism 41
3.3.1.2 The First-Person Perspective and Anti-Scientism 43
3.3.2 The Method of Heterophenomenology 47
3.3.2.1 What Is Heterophenomenology 48
3.3.2.2 Changing the Explanandum 50
3.3.2.3 Dennett's Official Reply 52
3.3.2.4 Phenomenal Realism 54

3.4 The Scope and Limits of this Project 55
3.5 Summary 57

4. Higher-Order Representation and Introspection 59
4.1 HOR, HOP, and HOT: Terminology 59
4.2 Is HOR A Theory of Qualitative Consciousness? 60
4.2.1 The Independence of HOR and Qualitative Consciousness 60
4.2.2 Armstrong's Absentminded Driver and the Explanandum of HOR 61
4.2.2.1 Reflections on Armstrong's Driver 62
4.2.3 HOR and What It's Like 68
4.3 Why Is HOR's Explanandum Elusive? 70
4.3.1 Qualia: Celluloid vs. the Silver Screen 70
4.3.1.1 Qualia and Secondary Qualities 73
4.3.1.2 The Reaction Criterion 74
4.3.2 Summary of the Analysis 75
4.4 HOR and Introspectionism 76
4.4.1 The Label 'Introspectionism' 76
4.4.2 Introspectionism: A Unitary Phenomenon? 79
4.4.3 Introspection: Perceptual or Purely Cognitive? 79
4.4.4 Is there an Introspective Phenomenology? 81
4.4.5 Perception without Phenomenology? 85
4.4.6 Narrowing the Gap between HOT and HOP 86
4.4.7 Externalizing Qualia 87
4.4.8 Qualia: Mental Paint vs. Physical Paint 89
4.5 Summary 91

5. The Allure of Introspectionism 93
5.1 Explaining Consciousness through Introspection: Introspectionism 93
5.2 Why Introspectionism Seems Obvious 94
5.2.1 The Definitional Link 94
5.2.2 The Best-Explanation Connection 95
5.2.3 Consciousness and the Consciousness of Consciousness 96
5.2.4 Introspectionism's Insight 97
5.2.5 Preserving the Insight 98
5.3 The Epistemology of Consciousness and Introspection 99
5.3.1 Three Forms of Epistemic Access to Our Own Minds 100
5.3.2 The Transition to Introspectionism: The Epistemic Argument 101
5.3.2.1 Identity: Introspections are Qualitative Beliefs 102
5.3.2.2 Presupposition: Qualitative Beliefs Presuppose Introspections 103
5.3.2.3 Special Epistemic Status: A Straw Man? 104

5.4 Severing Consciousness from Introspection 105
5.5 The Introspective Access Assumption 106
5.6 Against the Three Epistemological Theses 110
5.7 The Special-Knowledge Thesis 110
5.7.1 The Argument From Materialism 111
5.7.2 The Argument from Conceptualization 112
5.7.3 The Argument from Deviant Reasons 112
5.7.4 Assessment of the Special-Knowledge Thesis 113
5.8 The Fallible-Knowledge Thesis 114
5.9 The Belief Thesis 115
5.9.1 Detached Qualitative Beliefs 116
5.9.2 Detached Qualia 117
5.9.2.1 The First Appeal to Phenomenology 118
5.9.2.2 Simple Minds 123
5.10 Summary 126

6. Oscar, the Unconscious Introspector: A Case Study 129
6.1 Pollock's Oscar as Imitation Man 130
6.2 Rational Functionalism and Consciousness 130
6.2.1 Linking Rationality and Consciousness 130
6.2.2 The Role of Consciousness in the Human Rational Architecture 131
6.3 How to Build Consciousness into Oscar 132
6.4 Doubting that Oscar Is Conscious 135
6.4.1 Perception without Sensation 135
6.4.2 The Lessons of Blindsight 137
6.4.3 Pollock's Analysis of Blindsight 137
6.4.4 Introspection and Learning 138
6.4.5 The Capacities of Blindsighted Subjects 139
6.4.6 Blindsight and Learning: A Dilemma 140
6.4.7 The Argument from Blindsight 141
6.4.8 An Objection to the Argument from Blindsight 142
6.4.9 Reply to the Objection 142
6.5 Should We Take the Possibility of Oscar's Blindsight Seriously? 144
6.5.1 Two Arguments for Taking Oscar's Blindsight Seriously 144
6.5.1.1 The Argument from Parsimony 144
6.5.1.2 The Argument from Mystery 144
6.5.2 An Objection: The Two Arguments Prove too Much 145
6.5.3 Two Arguments Against Taking Oscar's Blindsight Seriously 146
6.5.3.1 The Argument From Analogy 146
6.5.3.2 The (Thought) Experimental Solution: Trying on Oscar 147

6.6 Summary 149

7. Relocating Qualia 151
7.1 Three (or Four) Ways of DeMenting Qualia 152
7.2 Denying Qualia 154
7.3 Wholehearted Relocation 156
7.3.1 The Expanding Sphere of Consciousness: The Inflation Problem 156
7.3.1.1 Rarefaction 158
7.4 Support for the Relocation Theses 161
7.4.1 The Second Appeal to Phenomenology 162
7.4.2 The Inconclusiveness of the Appeal to Phenomenology 163
7.4.3 No Mental Home for Qualia 167
7.5 Halfhearted Relocation: The Colored-Brain Thesis 168
7.5.1 Mental Representations Old and New 170
7.5.2 Qualia—"Nowhere to Be Found in the Head" 171
7.5.3 Exemplifying Qualia and Having Qualia 173
7.5.4 The Prospects of Halfhearted Relocation 174
7.6 The Price of Wholehearted Relocation 174
7.6.1 Does Wholehearted Relocation support the Belief Thesis? 175
7.6.2 Does Wholehearted Relocation undermine the Belief Thesis? 176
7.6.3 Wholehearted Relocation and the Existence of Qualia 177
7.7 Summary 179

8. Having Relocated Qualia 181
8.1 Complicating the Picture: Four Representation-Based Theories 182
8.2 Cell (1): Internal Qualia with HO Representation 185
8.2.1 The Problem of the Rock 185
8.2.2 Mediate and Immediate HO Thoughts 187
8.2.3 The Official Reply to the Rock 187
8.2.4 The Rock and Intrinsicalism 188
8.2.5 HOT: Analysis not Explanation 190
8.2.6 Assessing the Rock 194
8.3 Cell (2): Internal Qualia without HO Representation 195
8.4 Cell (3): External Qualia with HO Representation 199
8.4.1 Introspectionism and Relocation 199
8.4.2 The Inner Sense and the External World 200
8.5 Cell (4): External Qualia without HO Representation 202
8.5.1 Precursors: Naive Realism and Moore 203
8.5.2 Dretske's Representationalism 205
8.5.3 What Does Representationalism Explain? 207

8.6 Summary 209
8.7 My Position: A Preview 211

9. Denying Relocated Qualia 214
9.1 The Argument from Illusion and Nonexistent Qualia-Bearers 215
9.1.1 The Route to Nonexistent Qualia-Bearers 215
9.1.2 Intentional Objects—Mere Manners of Speech? 219
9.2 Problems With Intentional Objects 220
9.2.1 Intentional Objects à la (Early) Brentano 222
9.2.2 Intentional Objects à la Meinong 223
9.2.3 Intentional Objects and Possible Worlds 225
9.2.3.1 Lewisian Realism 226
9.2.3.2 Ersatzism 226
9.2.4 How to Perceive Otherworldly Objects 229
9.2.5 Halfhearted Relocationism and the Problems of Illusion 231
9.3 The Argument from Atomism and Illusory Qualia 232
9.3.1 Direct Reductive Realism about Color 234
9.3.2 The Illusion of Concrete Secondary Quality: Armstrong's Account 235
9.3.3 Piercing the Illusion 238
9.3.4 The Nature of the Illusion 239
9.3.4.1 Seeming and Believing 240
9.3.4.2 Seeming and the Language of Thought: Lycan's Way Out 242
9.3.5 The Origin of the Illusion 247
9.3.5.1 The Gestalt *Model* 247
9.3.5.2 Resemblances and Intrinsic Natures 248
9.3.5.3 The Headless-Woman Illusion 249
9.3.5.4 Should We Trust the Impression of Simplicity? 249
9.3.5.5 'Illusion' of Concrete Secondary Quality? 252
9.3.6 The Depth of the Illusion 253
9.3.6.1 First Interpretation: Total Illusion 254
9.3.6.2 Second Interpretation: Partial Illusion 255
9.3.6.3 Third Interpretation: Minimal Illusion 256
9.3.7 DIRE Realism and Illusory Colors 258
9.4 Summary 259

10. Consciousness: The Having of Qualia 262
10.1 Learning from Mistakes 262
10.2 Isolating Two Pernicious Assumptions 263
10.3 Adverbialism 265
10.3.1 Problems with Adverbialism 266

10.3.2 The Reappearance of the Two Pernicious Assumptions 267
10.3.3 The Phenomenal Adequacy of Adverbialism 271
10.4 Monism 273
10.5 Monism and the Having-Problem 273
10.5.1 The Neutral Theory of Experience 273
10.5.2 The Bundle Theory of the Self 276
10.5.3 The Phenomenal Adequacy of Monism 277
10.5.4 Challenging the Monistic Account of Having 279
10.5.5 Challenge One: Monism Is Unilluminating 279
10.5.6 Challenge Two: Monism Entails Infallibilism 280
10.5.6.1 The Noncognitive Nature of Monism 281
10.5.7 Challenge Three: Monism Entails Nonnaturalism 283
10.6 Monism and the Nature-Problem 284
10.6.1 The Bundle Theory of Percepts 285
10.6.2 The Physical Location of Percepts 288
10.6.2.1 The Colored-Brain Problem 292
10.6.3 The Metaphysical Location of Percepts 294
10.6.3.1 The Other Identity Theory 294
10.6.3.2 Matter as All Gap: Structural Realism 297
10.6.3.3 Filling the Gap: Qualia as the Basic Reality 298
10.6.3.4 Events, Matter, and Grain 301
10.6.4 The Threat of Panpsychism 304
10.7 Summary 307

Notes 311

References 332

Detailed Table of Contents 346

Index 352

Index

absence thesis, 154–56
    introduced, 152
Ackerman, Felicia, 318
acquaintance, 106–9. *See* introspective access assumption
    and knowledge by description, 108–9
act/object assumption, 264–65
    and self-presentation, 275
    and the Humean dissolution of the self, 274–75
    and the neutral theory of experience, 273–74
Adams, Douglas, 325
adverbialism, 265–73
    a Chisholmian argument against, 267–71
    and intentional objects, 220
    and qualia-bearing objects, 285
    and the act/object assumption, 264, 267–71
    and the individuation of species of sensing, 271–73
    and the recategorization of qualia, 266, 271
    and the substance/attribute assumption, 264, 267–71
    as a nonrelational theory of experience, 273–74
    the phenomenal adequacy of, 271–73
    the place of qualia in, 266, 271, 327
Alston, William
    and the first appeal to phenomenology, 119
    and the reach of introspection, 201
    on perception without phenomenology, 85–86
    on the range of qualia, 25
analysis
    vs. explanation. *See under* explanation
appearing. *See* seeming
argument from atomism. *See also* DIRE realism; illusion of concrete secondary quality
    and DIRE realism, 234–35
    and event ontology, 303–4
    and silver-screen qualia, 283–84, 292
    its connection with the argument from illusion, 258–59
    relevance of silver-screen qualia, 294–304
    stated, 232–33
argument from blindsight. *See* blindsight, argument from
argument from illusion, 215
    and nonexistent qualia-bearers, 215–31. *See also* intentional objects
Armstrong, David, ix, 311, 317, 326, 327
    and the long-distance driver, 61–68
    on DIRE realism, 234–39, 314
    on introspection, 111–12, 201
    on nonexisting intentional objects, 217
    on perception without phenomenology, 85–86, 135–36
    on perceptual consciousness, 65
    on phenomenal adequacy, 314, 321
    on phenomenological adequacy, 37
    on resemblance and intrinsic quality, 237, 248–49
    on secondary qualities, 234–35. *See also* DIRE realism
    on seeming as believing, 240–42
    on the externalization of qualia, 163

on the *Gestalt* model, 236, 247–48, 257
on the headless-woman illusion, 237, 249–52
on the identity theory, 294–96
on the illusion of concrete secondary quality, 235–39
on the importance of consciousness, 316
on the myopia of introspection, 325
on the special-knowledge thesis, 111–12
on the topic-neutral maneuver, 236–37
on the transparency of the mental, 159
atomism. *See* argument from atomism
Aune, Bruce, 326, 331
on rejecting qualia eliminativism, 3
Ayer, A.J., 317

Baldwin, Thomas, 323
Barnes, Jonathan, 2
behaviorism, 37–38
belief thesis, 100
and detached qualia, 117–23
and detached qualitative beliefs, 115–17
and the epistemic argument for introspectionism. *See* epistemic argument for introspectionism
and the relocation of qualia, 175–77
critique of, 115–26
detached qualia
argument from simple minds, 123–26
first appeal to phenomenology, 118–23, 176–77, 319
versions of, 115–16
versions of the, 319
Benham disk, 255
Berkeley, George, 254, 318, 323
Bisiach, E., 315
blindsight, 63, 64–66, 136–39
and perceptual consciousness, 64–66
and the capacities of blindsighted subjects, 139
in Oscar, 141–49
Pollock's analysis of, 136–37
the argument from, 141–44
blindsight, the argument from, 322
Block, Ned, 311, 319
on phenomenal and access consciousness, 317–18
on qualitative consciousness, 16
on the best explanation argument for consciousness, 146
Boghossian, Paul, 323
Brandl, Hannes, ix
Brentano, Franz, 326
and reism, 267
on intentional objects, 222
Broad, C.D., 107, 317, 326
bundle theory, 287–88, 327
of percepts, 285–88
and adverbialism, 287
and the exemplification problem, 285–87
and the substance attribute assumption, 287
of the self, 276–77
and self-presentation, 276–77
and the monistic account of having, 276–77
Butchvarov, Panayot
on adverbialism, 265, 270
on science vs. philosophy, 43
on the philosophical investigation of consciousness, 45

Campbell, Keith, 130
Carruthers, Peter, 311
on animal consciousness, 320
Castañeda, Hector-Neri, 110
on the nature of philosophy, 55
causal theory of percpetion. *See under* percpetion
causation problem
and ersatzism about possible worlds, 228
as part of the having-problem, 221
for homeless objects, 224–25

for Lewisian realism, 226
for otherworldly objects, 229–31
for sense-data, 229
introduced, 221
celluloid qualia, 70–72
and the relocation of qualia, 153
criterion for, 75
introduced, 71
Chalmers, David, ix, 330
on easy vs. hard problems of consciousness, 318
Chisholm, Roderick, 144, 318, 326, 328
and adverbialism, 267–71, 327
on Brentano's reism, 267
on homeless objects, 223
on intentional inexistence, 222–23
Chuang-Tze, 126
Churchland, Paul, 144
on perception without phenomenology, 136
cognitive science. *See* methodology
colored brain thesis. *See* HRT
colored-brain problem, 171–73, 291–94
stated, 30
conscious inessentialism, 18, 136
Own Flanagan's definition of, 18
consciousness. *See also* qualitative consciousness
and content, 313
confused with the knowledge of consciousness. *See* introspectionism; epistemic argument for introspectionism
fully introspective, 62
intransitive, 187, 191–92
introspective vs. nonintrospective, 77, 316
methods of studying. *See* methodology
noncognitive nature of, 281–83
of animals. *See* simple minds
other notions of, 16–18
perceptual. *See* perceptual consciousness
phenomenal vs. access, 317–18
qualitative. *See* qualitative consciousness
state. *See* state consciousness
theories of
adverbialism. *See* adverbialism
hierarchical. *See* HOR; introspectionism
higher-order. *See* HOR; introspectionism
higher-order experience. *See* HOP
higher-order perception. *See* HOP
higher-order representation. *See* HOR; introspectionism
higher-order thought. *See* HOT
intentionalism. *See* representationalism
introspectionism. *See* introspectionism; HOR
minds-just-do-it theory. *See* minds-just-do-it theory of consciousness
monism. *See* monism
noncognitive. *See* monism
representationalism. *See* representationalism
representation-based theories. See representation-based theories
representation-use theory. *See* representation-use theory of consciousness
transitive, 187, 191–92
constant cognition, thesis of, 101
Crane, Tim, 222
Crick, Francis
on the neurophysiological correlates of qualia, 196
Cummins, Robert, ix, 210
on bruteness of explanation, 280
on the strangeness of consciousness, 321
on trying on Oscar, 147–49

David, Marian, ix
Delaney, Cornelius, 326

deMenting qualia, 152–54. *See also* WRT, HRT, absence thesis
Democritus of Abdera, 2, 232
Dennett, Daniel, ix, 30, 311, 322
  and heterophenomenology, 47–55. *See also* heterophenomenology
  and materialism, 315
  and qualia eliminativism, 21–24, 154–56, 312
  and the absence thesis, 154–56
  and the defense of the third-person perspective, 47–55
  and the first-person perspective, 40, 44
  first-person operationalism, 23–24
  on elimination vs. reduction, 322–23
  on illata and abstracta, 315
  on science vs. philosophy, 315
  on seeming and judging, 52–54. *See also* seeming
  on the exemplification/having distinction, 173
detached qualia. *See under* qualia
detached qualitative beliefs. *See under* epistemic argument for introspectionism, qualitative belief
diaphanousness of the mental. *See* qualitative consciouosness, rarefaction of
DIRE realism (direct reductive realism about color), 234–39
  and the argument from atomism, 234–35
  and the illusion of concrete secondary quality, 235–39. *See also* illusion of concrete secondary quality
  and the simplicity of qualia, 240–47
  phenomenal adequacy of, 254
direct reductive realism (about color). *See* DIRE realism
Dretske, Fred, 320, 323, 324
  and the colored brain problem, 197, 292–93
  and the first appeal to phenomenology, 119
  his notion of qualitative consciousness critized, 207–9
  on qualia, 206
  on representationalism, 205–9
  on systemic representation, 206
  on the dispute between HOT and HOP, 86–88
  on the externalization of qualia, 157
  on the myopia of introspection, 201
  on the phenomenology of introspection, 83

Eddington, Arthur, 232
Edgell, Beatrice, 107
Einstein, Albert, 311
  on the importance of qualitative consciousness, 1, 23
eliminativism of qualia. *See under* qualia
epistemic argument for introspectionism, 100–105. *See also* introspectionism
  and the introspective access assumption. *See under* introspectionism
  based on a confusion of levels, 96–97
  belief thesis. *See* belief thesis
  critique of, 106–25
  fallible-knowledge thesis, 100
    critique of, 114–15
  special-knowledge thesis, 100
    critique of, 111–14
  stated, 103
  thesis of constant cognition, 101
event ontology, 301–4
  and the other identity theory, 302–3
  and the problem of atomism, 303–4
  examples of events, 302
  physical particles constructed from events, 302
exemplification problem
  and ersatzism about possible worlds, 227

and the bundle theory of percepts, 285–87
and the substance/attribute assumption, 285
as part of the nature-problem, 221
introduced, 221
exemplification/having distinction for qualia, 173–74, 197–98
experience, neutral theory of. *See* neutral theory of experience
explanation
bruteness of, 280
informative, 185
vs. analysis, 1–2, 185, 190–94
explanatory gap, problem of, 243
externalism
of content, 156–58
of qualia, 87–89, 157–58, 177–79. *See also* WRT; physical paint
second appeal to phenomenology, 162–67

fallible-knowledge thesis. *See under* epistemic argument for introspectionism
Farell, B.A., 68, 311
feeling
and knowing, 282–83
as noncognitive, 281–83
Feigl, Herbert, 20, 311, 330
on appearance vs. reality, 299–300
on evidence and reference, 299–300
on panpsychism, 304–6
on qualia as basic reality, 298–301
on the other identity theory, 297, 299–300
first appeal to phenomenology. *See under* belief thesis, detached qualia
first-person operationalism, 23–24
first-person perspective, 38–55
and heterophenomenology, 47–55
and materialism, 41–42
and the scientific perspective, 43
Dennett's critique of, 40
Flanagan, Owen, 280, 315
on conscious inessentialism, 18
on new mysterianism, 47
on qualia eliminativism, 312
on the range of qualia, 25
Fodor, Jerry, 300, 304
and materialism, 31
on bruteness of explanation, 280
functionalism, rational, 130–32

Geach, Peter, 25
Gestalt model, 257. *See under* illusion of concrete secondary quality, the origins of
Globus, Gordon, ix
Goldman, Alvin, 311, 315
on consciousness and content, 313
on the problem of the rock, 186
on the range of qualia, 25
Graham, George, 178
grain argument. *See* argument from atomism
Griffin, Donald
on simple minds, 126
Gunderson, Keith, 144
on the prospects of AI, 321
Güzeldere, Güven
on the dispute between HOT and HOP, 86–88

Haller, Rudolf, ix
Hannay, Alastair
on the what it's like locution, 14
Harman, Gilbert, 89
on nonexisting intentional objects, 216
on the externalization of qualia, 163
Haugeland, John, 131
having-problem, 4–5, 151, 262–64
and monism, 277
and noncognitivism, 281–83
and the act/object assumption, 263
and the causation problem, 221. *See also* causation problem
attempted solutions of. *See under* consciousness, theories of; qualitative consciousness, theories of

detached qualia, their relevance for, 176
first appeal to phenomenology, its relevance for, 177, 320
for relocated qualia, 181–210
HOR as theory of, 70
monistic answer to, 277
as unilluminating, 279–80
stated, 33–34
headless-woman illusion. See under illusion of concrete secondary quality, the origins of
Helen (the blindsighted rhesus monkey), 139
Hess, Peter, 119
heterophenomenology, 47–55
and the distinction between seeming and judging, 53. *See also* phenomenal realism
and the third-person perspective, 47–55
Daniel Dennett's defense of, 47
datum, heterophenomenological, 50
object, heterophenomenological, 49
world, heterophenomenological, 48
Hicks, Dawes, 107
hierarchical theory of consciousness. *See* introspectionism; HOR
higher-order experience theory of consciousness. *See* HOP. *See also* introspectionism
higher-order perception theory of consciousness. *See* HOP; introspectionism
higher-order theory of consciousness. *See* HOR; introspectionism
higher-order thought theory of consciousness. *See* HOT; introspectionism
Hilbert, David, 326
Hofstadter, Douglas, 311
Holt, Edwin B., 158
homeless objects. *See under* intentional objects
homogeneity (of qualia). *See* qualia, simplicity and homogeneity of
HOP (higher-order perception theory of consciousness). *See* HOR; introspectionism
and the problem of the rock. *See* rock, the problem of the
augmented with the language of thought, 242–47
vs. HOT, 79–87, 317
HOR (higher-order representation theory of consciousness). *See also* introspectionism
and introspectionism, 76–79
and perceptual consciousness, 64–66
and qualitative consciousness, 60–70
and the having-problem, 70
and the problem of the rock. *See* rock, the problem of the
vs. HOT, 317
HOT (higher-order thought theory of consciousness). *See* HOR; introspectionism
and direct higher-order thoughts, 187
and the problem of the rock. *See* rock, the problem of the
explanation or analysis, 190–94
vs. HOP, 79–87
HRT (halfhearted relocation thesis). *See under* relocation theses
Hume, David, 32
on acquaintance with the self, 248
on the dissolution of the self, 274–75
Humphrey, Nicholas
on blindsight, 139
on sensation vs. perception, 321
on simple minds, 125
on the importance of qualitative consciousness, 57
identity theory, 294–96, 330. *See also* other identity theory
and qualia, 295–96
and the topic neutrality about the mind, 295–96
illusion of concrete secondary quality, 235–59
and the Muller-Lyer illusion, 252

and the sense-datum theory, 239–40
depth of, 252–58
minimal illusion, 256–58
partial illusion, 255–56
total illusion, 254–55
not really an illusion, 252–53
the nature of, 239–40
doxastic account, 240–42
language of thought account, 242–47
the origins of
Gestalt model, 236, 247–48
headless-woman illusion, 237
healess woman illusion, 249–52
resemblance illusion, 237, 248–49
imitation man, 130
infallibilism (about qualitative mental staes), 280–81
infallibilism (about qualitative mental states
and noncognitivism, 281–83
inflation problem, 156–61. *See* WRT; externalization of qualia
and the myopia of introspection, 202
inner sense. *See* introspection
intentional objects, 215–31
Alexius Meinong's conception of, 223–25
and possible worlds, 225–31
and the causation problem, 221. *See also* causation problem
and the exemplification problem, 221. *See also* exemplification problem
as homeless objects, 223–25
and possible worlds, 224
and the exemplification problem, 224
as mere manners of speech, 219–20
constrained by the WRT, 220–21
Franz Brentano's conception of, 222
nonexisting, 215–19
as bearers of qualia, 215–19
nonphysical, 218
intentionalism. *See* representationalism
intransitive consciusness. *See under* consciousness
intrinsicalism (about qualitative consciousness)
and relational accounts of consciousness, 188–90
and the subjectivity of consciousness, 188–90, 194. *See also* subjectivity
introspection
and cognition of the mental, 317
and learning, 138
and myopia, 200–201, 323–24
as inner sensing, 79–86
as neutral between HOT and HOP, 77–78
distinguished from qualitative beliefs, 102
hierarchical vs. nonhierarchical accounts, 78, 316, 320
its access to the external world, 200–202
John Locke on, 77
privileged epistemic status of, 110–14
the phenomenology of, 79–86. *See also* perception
introspectionism. *See also* HOR
and acquaintance, 318
and myopia, 200–201
and the introspective access assumption, 106–10
and acquaintance, 106–9
stated, 106
and the problem of the rock. *See* rock, the problem of the
and the relocation of qualia, 174–77
and the subjectivity of consciousness, 97–100. *See also* subjectivity
argument from blindsight, 141–44, 322
argument from simple minds. *See under* belief thesis

as an account of the having-problem, 70
as the common core of HOT and HOP, 76–79
based on a level confusion, 96–97. *See also* epistemic argument for introspectionism
best explanation argument for, 95–96
combined with the WRT, 324, 199–200
definition-based argument for, 94–95
does not address the nature-problem, 70
epistemic argument for. *See* epistemic argument for introspectionism
first appeal to phenomenology. *See under* belief thesis
introduced, 76
Oscar as incarnation of, 132–35
supported by cognitive intertia, 105
introspective access assumption. *See under* introspection

James, William, 20, 305, 328
Jaynes, Julian, 96

Kihlstrom, John, 144
Kim, Jaegwon, 158, 323
knowing
and feeling, 282–83
as relational, 282–83
knowledge argument, 331

Lackey, Jennifer, ix
Landesman, Charles, 3
Lashley, Karl, 132
learning
and blindsight, 139–41
and introspection, 138
Lehrer, Keith, ix, 326
on inrospection, 112–13
on the special-knowledge thesis, 112–13, 319
level confusion (about consciousness), 96–97
Lewis, David, 226
Lloyd, Dan
on level confusions about consciousness, 96
on simple minds, 125
Locke, John
and his definition of consciousness, 94–95
on introspection, 77
on the Mentality of qualities, 2–3
Lockwood, Michael, 314
on dualism, 330
qualitative consciousness, 27
Lycan, William, ix, 311, 317, 322, 325, 326, 327
and fully introspective consciousness, 62
and the problem of the rock, 192–93
his notion of a quale, 88
on HOR as theory of qualitative consciousness, 60–61, 68
on impure functionalism, 325
on nonexisting intentional objects, 217–18
on perceptual consciousness, 64–65
on qualia eliminativism, 36–37
on the appearance and reality of qualia, 73
on the causation problem for otherworldly objects, 229–30
on the causation problem for sense-data, 229
on the explanatory gap, 243
on the externalization of qualia, 88
on the language of thought account of seeming, 242–47
on the phenomenology of introspection, 80, 82–83
on the stereoscopic fallacy, 28–29
Lyons, William, 316
on the phenomenology of introspection, 82–83

Mach, Ernst, 305, 328

Makie, J.L.
on intentional objects, 219–20
Malebranche, Nicolas, 320
Maloney, J. Christopher, ix, 324
and representationalism, 205
on exemplifying and having qualia, 197–98
on the nature of qualia, 196
on the representation-use theory of consciousness, 195–99
Marcel, A.J., 315
Marek, Hannes, ix
materialism
and nonphysical merely intentional objects, 218
and qualia, 31
and qualitative consciousness, 31
and the first-person perspective, 41–42
vs. physicalism, 314
vs. science, 314
Maudlin, Tim
on qualitative consciousness, 27
on the strangeness of consciousness, 322
Maxwell, Grover, 330, 331
on structural realism, 298
on the location of percepts, 290
McGinn, Colin
and materialism, 41
and new mysterianism, 46
and the first-person perspective, 41
on causation, 225
on content externalism, 157
on qualitative consciousness, 26–27
Meinong, Alexius
on intentional objects as homeless objects, 223–25
Mellor, H.D., 321
mental paint qualia, 89–90. *See also* HRT
and the veil of perception, 166
how they would be experienced, 166–67
Mental/mental distinction, 311
and the HRT, 169
merely intentional objects. *See* intentional objects, nonexisting
Merricks, Trenton, ix
methodology (of studying consciousness)
behaviorism, 37–38
cognitive science, 37–38
first-person perspective. *See* first-person perspective
heterophenomenology. *See* heterophenomenology
new mysterianism, 46–47
scientific vs. philosophic, 1–2, 57, 43–55, 315
third-person perspective. *See* third-person perspective
minds-just-do-it theory of consciousness, 204, 225, 230, 278
monism, 273–310
and event ontology, 301–4
and infallibilism, 280–81
and naturalism, 283–84
and noncognitivism, 281–83
and nonnatural qualia, 283–84
and panpsychism, 304–7. *See also* panpsychism
and percepts. *See* percepts
and qualia as basic reality, 298–301
and structural realism. *See* structural realism
and the act/object assumption, 265
and the bundle theory of percepts. *See* bundle theory, of percepts; percepts
and the bundle theory of the self. *See* bundle theory, of the self
and the colored brain problem, 291–94
and the having-problem, 273–84
and the neutral theory of experience. *See* neutral theory of experience
and the other identity theory. See other identity theory
and the substance/attribute assumption, 265

challenged as unilluminating, 279–80
having-problem
monistic answer stated, 277
nature-problem, 284–307
and event ontology, 301–4
monistic answer stated, 301
phenomenological adequacy of, 277–80
previewed, 98, 211–13
summarized, 307–10
Moore, G.E., 107
and naive realism, 203–5
and the 'minds do it' theory of having qualia, 204
on direct awareness, 204
on sense-data, 21, 170, 312
on the transparency of consciousness, 159–61
Moravec, Hans, 323
Moser, Paul, 320
Muller-Lyer illusion, 252
myopia, the problem of, 325
and myopia, 323–24
and the inflation of the mind, 202
introspection affected by, 200–201

Nagel, Thomas, 13, 131, 145, 311, 322, 323
and materialism, 41
and new mysterianism, 46
and physicalism, 42, 45
and the first-person perspective, 41
on the exemplification/having distinction, 173
on the reduction of secondary qualities, 73
qualitative consciousness, 13–14
what it's like locution, 13
naive realism, 203–5. *See also* representationalism
and direct awareness, 203–4
nature-problem, 4–5, 151, 262–64
and event ontology, 301–4
and the exemplification problem, 221. *See also* exemplification problem
and the substance/attribute assumption, 263
detached qualia, their irrelevance for, 176
first appeal to phenomenology, its irrelevance for, 177, 320
monistic answer to, 301
not addressed by HOR, 70
stated, 33–34
neon theory of consciousness. *See* monism
neutral monism, 297, 328
and panpsychism, 305
neutral theory of experience, 273–76
and the act/object assumption, 273–74
and the Humean dissolution of the self, 274–75
and the location of percepts, 288
neutralism. *See* neutral theory of experience
neutrality thesis, 19–20
and phenomenological adequacy, 37
new mysterianism, 46–47
Nisbett, R.E., 326
nonrelational theory of experience. *See* neutral theory of exeprience; adverbialism

Oscar (John Pollock's artificial person), 129–50. *See* HOR. *See also* introspectionism
and blindsight, 141–49
and learning, 138
and the argument from blindsight, 141–44, 322
and the trying on test, 147–49
as imitation man, 130
as introspective machine, 132–35
introduced, 130
other identity theory, 296–304. *See also* identity theory
and event ontology, 301–4
and panpsychism, 304–7
and qualia. *See* qualia, as basic reality

and structural realism. *See* structural realism
and topic neutrality about matter, 296–98

panpsychism, 304–7
absurdity of, 307
and neutral monism, 305
vs. materialism, 306
vs. physicalism, 306
Parsons, Terence, 313
perception
and HOT vs. HOP, 85–86
causal theory of, 290–91
and the location of percepts, 290–91
of one's own brain. *See* percepts, location of
without phenomenology, 85–86, 135–38. *See also* blindsight
percepts
and the colored-brain problem. *See* colored-brain problem
and the substance/attribute assumption, 285–86
expelled from the world of physics, 289–90, 329–30
introduced, 275
location of, 288–92
and the causal theory of perception, 290–91
and the colored-brain problem, 291–94
and the neutral theory of experience, 288
and the privacy of one's perceptual world, 291, 297
in the brain, 288–92
metaphysical status of, 294–304
self-presenting, 275. *See also* self-presentation
the bundle theory of, 285–88
perceptual consciousness, 65–66
and HOR, 64–66
distinguished from blindsight, 53–55
phenomenal object. *See* intentional object; percepts
phenomenal properties. *See* qualia
phenomenal realism, 53–55
phenomenalism, 330
phenomenological adequacy
and the doxastic account of seeming, 240–42
and the first-person perspective. *See* first-person perspective
and the language of thought account of seeming, 242–47
and the neutrality of qualia, 37
and the trying on test, 38
and trying on Oscar, 147–49
of adverbialism, 271–73
of DIRE realism, 254, 314
of monism, 277–80
principle of, 36–38
Searle, John, 38
philosophy vs. science. *See* methodology
physical paint qualia, 89–90. *See* externalism, of qualia; qualia, externalization of; WRT
Place, U.T., 159
Planck, Max, 289
Plantinga, Alvin
and the first appeal to phenomenology, 119
on possible worlds, 226–27
on the range of qualia, 25
Pollock, John, ix, 319, 326
and the first appeal to phenomenology, 119
and the Oscar project, 129–50
on blindsight, 63, 66, 136–37
on introspectoin, 113
on qualitative states, 143
on rational functionalism, 130–32
on sensation, 316, 321
on the special-knowledge thesis, 113
possible worlds, 226
and merely intentional objects, 225–29
ersatzism about, 226–29

Lewsian realism about, 226
Price, H.H., 224, 323
on the neutrality thesis, 312
property instantiation/exemplification
and the bundle theory, 287–88
Prutti, Brigitte, ix
Putnam, Hilary, 157

qualia
and anti-physicalism, 152
and sense-data, 21
and the language of thought, 196
appearance vs. reality, 73, 283
as basic reality, 298
and panpsychism, 304–7
as mental states or events, 312
celluloid. *See* celluloid qualia
Christopher Maloney's notion of, 196
deMenting of, 152–54. *See* relocation theses; absence thesis
detached (from qualitative beliefs), 117–23
and the relocation of qualia, 175–77
and the WRT, 176
elimination of, 3, 21–24, 36–37, 312
epistemic problems raised by, 313
exemplification vs. having, 173–74
expelled from the world of physics, 2–3, 289–90, 329–30
externalization of, 87–89, 157–58, 177–79. *See also* WRT
second appeal to phenomenology, 162–67
Fred Dretske's notion of, 206
importance of, 24–26
individuation of, 22–24
introduced & defended, 18–32
mental paint. *See* mental paint qualia
naturalizability of, 6, 283–84
neurophysiological correlates of, 196
neutrality of. *See* neutrality thesis
not borne by Mental objects, 167–68
not necessarily in the mind, 177–79. *See also* neutrality thesis
physical paint. *See* physical paint qualia
range of, 25–26
recategorized in adverbialism, 266
reduction of. *See* DIRE realism
to mental paint. *See* HRT
to physical paint. *See* WRT
relocation of. *See* relocation theses
and celluloid vs. silver-screen qualia, 153
silverscreen. *See* silver-screen qualia
simplicity and homogeneity of. *See also* illusion of concrete secondary quality
and event ontology, 303–4
apparent or real, 249–52
as problem for DIRE realism, 240–42
doxastic account of, 240
language of thought account of, 242–47
strangeness of, 26–27
William Lycan's notion of, 88
qualia-light. *See* celluloid qualia
qualitative belief
as presupposing introspection, 103
containment thesis, 115
detached, 115–17
distinguished from the acts of introspection, 102
functional identity thesis, 115
identity thesis, 115
its transparent character, 101
qualitative consciousness
and content, 313
and its place in the philosophy of mind, 55–57
and materialism, 31–32
and the long-distance driver, 61–68
epistemic problems raised by, 313
feeling vs. knowing, 281–83
Fred Dretske's notion of, 207–9

importance of. *See* qualia, importance of
intrinsicalism about. *See* intrinsicalism (about qualitative consciousness)
introduced, 13–17
its subjectivity, 97–100. *See also* subjectivity
methods of studying. *See* methodology
Ned Block's notion of, 16
noncognitive nature of, 281–83
range of. *See* qualia, range of
rarefaction of, 158–61
strangeness of. *See* qualia, strangeness of
theories of
adverbialism. *See* adverbialism
hierarchical. *See* HOR; introspectionism
higher-order. *See* HOR; introspectionism
higher-order experience. *See* HOP
higher-order perception. *See* HOP
higher-order representation. *See* HOR; introspectionism
higher-order thought. *See* HOT
intentionalism. *See* representationalism
introspectionism. *See* introspectionism; HOR
minds-just-do-it theory. *See* minds-just-do-it theory of consciousness
monism. *See* monism
noncognitive. *See* monism
representationalism. *See* representationalism
representation-based theories. See representation-based theories
representation-use theory. *See* representation-use theory of consciousness
Thomas Nagel's notion of, 13
what it's like locution, 13–15
Quine, W.V., 56

Radner, Daisie, 320
Radner, Michael, 320
Ramsey, William, ix
raw feels. *See* qualia
realism
direct reductive. *See* DIRE realism
reductionism
direct realistic. *See* DIRE realism
Reid, Thomas
on feeling, 328–29
on introspection, 126
Relational accounts of consciousness. *See* intrinsicalism
relocation of qualia
ambiguity of, 183
relocation theses
and celluloid vs. silver-screen qualia, 153
and the belief thesis, 175–77
and the critique of introspectionism, 174–77
HRT (halfhearted relocation thesis), 168–73
and nonexistent qualia bearers, 231
and problems created by nonveridical experience, 231
and sense-datum theory, 170–71
and the colored brain problem. *See* colored brain problem
and the duplication of color properties, 172
and the Mental/mental distinction, 168
introduced, 152
support for, 161–68
WRT (wholehearted relocation thesis), 156
and content externalism, 156
and detached qualia, 176
and myopia, 200–202, 323–24

and nonexistent qualia bearers, 215–31. *See also* intentional objects
and the existence of qualia, 177–79
and the externalization of qualia. *See* qualia, externalization of
and the first appeal to phenomenology, 176–77
combined with introspectionism, 199–200, 324
introduced, 152
problems created by nonveridical experience, 215–31
second appeal to phenomenology, 162–67
physical vs. mental paint, 163–67
representation
ambiguity of, 182–83
as qualia-bearing object, 182–83
as qualia-revealing act, 182–83
systemic, 206
representationalism, 202–9
and its intended explanandum, 207–9
and naive realism, 203–5
and systemic representation, 206
and the representation-use theory of consciousness, 205
Dretske on, 205–9
representation-based theories of consciousness. *See* HOP; HOT; HOR; introspectoinism; representation-use theory of consciousness; representatoinalism; intentionalism
critique leveled against all versions of, 209–10
table of the four types of, 184
representation-use theory of consciousness, 195–99
and the colored brain problem, 196–97
and the language of thought, 196
Robinson, Howard, 314
Robinson, William, 323
on the diaphanousness of consciousness, 161
rock, the problem of the, 185–95
and intrinsicalism about consciousness, 188–90
David Rosenthal's reply, 187–88
Lycan's reply, 192–93
stated, 186
Rogers, Robert, ix
Rorty, Richard, 73
Rosenthal, David, 317, 324
and intrinsicalism about consciousness, 188–90
and the argument from reporting and expressing, 324
fully introspective consciousness, 62
his reply to the problem of the rock, 187–88
on celluloid qualia, 316
on explanation vs. analysis, 185, 190–94
on HOR as theory of qualitative consciousness, 60–61, 68
on perceptual consciousness, 66
on the appearance and reality of qualia, 73–74
on the externalization of qualia, 88–89
on the myopia of introspectoin, 325
on the phenomenology of introspection, 80–81
on the problem of illusion for the HRT, 231
on the unmediated nature of higher-order thoughts, 187
state consciousness, 15
Russell, Bertrand, 6, 322, 328, 330
and neutral monism, 297, 305, 328
and panpsychism, 305–6
and the act/object distinction, 328
and the knowledge argument, 331
and tropes, 329
on acquaintance, 106–9, 318
on consciousness and content, 313
on event ontology, 301–4

on giving up the self, 274
on knowledge, 282–83
on physicalism, 314
on sense-data, 171, 312
on space, 330
on the bundle theory of percepts, 286
on the causal theory of perception, 290–91
on the identity of sensing and sensed, 275
on the location of percepts, 288–92
on the other identity theory, 297
on the range of qualia, 25
on the rewards of philosophical analysis, 9
Rutte, Heiner, ix

Sanford, David, 316
Sayre, Ken, ix
Schlick, Moritz, 330
and noncognitivism, 281
on knowledge, 282–83
on the identity of sensing and sensed, 283
on the identity of sensing and the sensed, 328
science vs. philosophy. *See* methodology
Searle, John, 324
and materialism, 41
and the first-person perspective, 41, 43–45
and the trying on test, 38
on phenomenological adequacy, 38
on the reduction of secondary qualities, 73
second appeal to phenomenology. *See under* relocation theses
secondary qualities
acquaintance with their intrinsic natures, 248–49
and qualia, 73–75
expelled from the world of physics, 2–3, 289–90, 329–30
reduction of, 73–75
topic-neutral account of, 236–37
seeming
and believing, 240–42
and judging, 52–54
and the language of thought, 242–47
doxastic vs. phenomenological, 242
self
abandonment of, 274–75
bundle theory of. *See* bundle theory, of the self
self-presentation, 275, 327. *See also* percepts
and infallibilism, 280–81
and nonnaturalism, 283–84
without a self, 277
Sellars, Wilfrid, 313, 317, 326
on the grain argument, 233
on the importance of qualia, 26
sense, inner. *See* introspection
sense-data, 21, 167–68, 170–71, 312
and intentional objects, 220
identified with sensings, 275, 283. See also percepts
Shoemaker, Sidney, 115
on qualitative beliefs, 318
on the phenomenology of introspection, 84
silver-screen qualia, 70–72
and grain problems. *See* argument form atomism
and the atomic nature of matter. *See* argument from atomism
and the relocation of qualia, 153
as nonnatural, 283–84
criterion for, 75
introduced, 71
simple minds, argument from. *See under* belief thesis
simplicity (of qualia). *See* qualia, simplicity and homogeneity of
Smart, J.J.C.
and materialism, 31
on the transparency of the mental, 159
special-knowledge thesis. *See under* epistemic argument for introspectionism

Sprigge, T.L.S., 15
Stamenov, Maxim, ix
state consciousness. *See under* consciousness
    Rosenthal, David, 15
    subject involvement, 15
Stephens, Lynn G., 178
Stereoscopic fallacy, 28–29
    and materialism, 29
Sternberg, Robert, 247
Stove, David, 323
    on 'in' as the source of all metaphysics, 314
structural realism, 297–301
    and event ontlogy, 301–4
    and external objects
        known only by causal role, 298
        unknown intrinsic natures, 298
    matter known only absractly, 299
    qualia as basic reality, 298–301
Stubenberg, Leopold, 314, 317, 330
Sturgeon, Scott, ix
subjectivity (of qualitative consciousness), 97–100, 194
    and relational account of consciousness, 188–90
    without introspection, 212–14
substance/attribute assumption, 264–65

thick internalism. *See* introspectionism, combined with the WRT
    introduced, 211
thin externalism. *See* representationalism
    introduced, 211
third-person perspective. *See also* first-person perspective
    and reductionism, 45
    and the method of heterophenomenology, 47–55
Tolliver, Joseph, ix
topic-neutral maneuver
    applied to secondary quality concepts, 236–37, 238
transitive consciousness. *See under* consciousness
transparency of the mental. *See* qualitative consciouosness, rarefaction of
trope theory, 329
trying on test, 38, 147–49. *See also* phenomenological adequacy
Tully, Robert, 330
Twardowski, Kasimir, 182
Tye, Michael
    and adverbialism, 327
    on phenomenal realism, 53
    on qualia externalism, 157

Velleman, David, 323

Warfield, Fritz, ix
Weiskrantz, Laurence, 321
    on blindsight, 139
what it's like locution, 13–15, 13
Whitehead, Alfred North, 6
Willson, T.D., 326
Wisdom, John, 21
Wittgenstein, Ludwig, 26, 163–64
WRT (wholehearted relocation thesis). *See under* relocation theses

Zihl, Josef, 139
Zimmerman, Dean, ix

In the series ADVANCES IN CONSCIOUSNESS RESEARCH (AiCR) the following titles have been published thus far or are scheduled for publication:

1. GLOBUS, Gordon G.: *The Postmodern Brain*. 1995.
2. ELLIS, Ralph D.: *Questioning Consciousness. The interplay of imagery, cognition, and emotion in the human brain*. 1995.
3. JIBU, Mari and Kunio YASUE: *Quantum Brain Dynamics and Consciousness. An introduction*. 1995.
4. HARDCASTLE, Valerie Gray: *Locating Consciousness*. 1995.
5. STUBENBERG, Leopold: *Consciousness and Qualia*. 1998.
6. GENNARO, Rocco J.: *Consciousness and Self-Consciousness. A defense of the higher-order thought theory of consciousness*. 1996.
7. MAC CORMAC, Earl and Maxim I. STAMENOV (eds): *Fractals of Brain, Fractals of Mind. In search of a symmetry bond*. 1996.
8. GROSSENBACHER, Peter G. (ed.): *Finding Consciousness in the Brain. A neuro-cognitive approach*. n.y.p.
9. Ó NUALLÁIN, Seán, Paul MC KEVITT and Eoghan MAC AOGÁIN (eds): *Two Sciences of Mind. Readings in cognitive science and consciousness*. 1997.
10. NEWTON, Natika: *Foundations of Understanding*. 1996.
11. PYLKKÖ, Pauli: *The Aconceptual Mind. Heideggerian themes in holistic naturalism*. 1998.
12. STAMENOV, Maxim I. (ed.): *Language Structure, Discourse and the Access to Consciousness*. 1997.

DATE DUE / DATE DE RETOUR

FEB 0 6 2001

NOV 2 9 2001

NOV 3 0 2001

APR 0 6 2002

APR 0 3 2002

DEC 2 0 2004

CARR McLEAN

38-297

TRENT UNIVERSITY
0 1164 0410583 9

Consciousness and Qualia is a philosophical study of qualitative consciousness, characteristic examples of which are pains, experienced colors, sounds, etc. Consciousness is analyzed as the *having* of *qualia*. Phenomenal properties or *qualia* are problematical because they lack appropriate bearers. The relation of *having* is problematical because none of the typical candidates for this relation — introspection, inner monitoring, higher level thoughts — explains why it is like something to have a quale. The *qualia problem* is solved by introducing a bundle theory of phenomenal objects. Phenomenal objects are bundles of qualia. Thus there is no need for independent qualia bearers. The *having problem* is solved by introducing a bundle theory of the self. To have a quale is for it to be in the bundle one is. Thus no further relations are needed to explain how qualia are had. This study strives for phenomenological adequacy and thus the first-person point of view dominates throughout.

JOHN BENJAMINS PUBLISHING COMPANY