INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9307666

Consciousness and qualia

Stubenberg, Leopold, Ph.D.

The University of Arizona, 1992

Copyright ©1992 by Stubenberg, Leopold. All rights reserved.

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

CONSCIOUSNESS AND QUALIA

by

Leopold Stubenberg

Copyright © Leopold Stubenberg 1992

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 9 2

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read
the dissertation prepared by Leopold Stubenberg
entitled CONSCIOUSNESS AND QUALIA

and recommend that it be accepted as fulfilling the dissertation requirement for
the Degree of Doctor of Philosophy

Keith Lehrer

11-2-92

Date

Christopher J. Maloney

11/02/92

Date

Joseph Tolliver

11-2-92

Date

Final approval and acceptance of this dissertation is contingent upon the
candidate's submission of the final copy of the dissertation to the Graduate
College.

I hereby certify that I have read this dissertation prepared under my direction
and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director Keith Lehrer

11-2-92

Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under the rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: _____

A handwritten signature in cursive script, appearing to read "Robert J. Schubert", written over a horizontal line.

ACKNOWLEDGMENTS

This dissertation has been long time in the making. The one who has helped me most is Marian David. For more years than I care to mention he has suffered through interminable qualia confusions with me. Even when he thought that my ideas were “quite mad,” he proved patient and illuminating. I fear that many of the thoughts expressed here belong to the “quite mad” category. He is not to be held accountable for them. For a while Scott Sturgeon and I shared uncannily similar views about the nature and importance of qualia. He has now moved on to a more scientifically enlightened view of the matter. But I am sure that he will find echoes of our qualia discussions on every page of this dissertation. Robert Cummins and I hardly could have more divergent views about how to do philosophy of mind. Nevertheless he has supported me and my philosophical interests over many years. I have learned much from him. I do hope this thesis will not disappoint him. I profited greatly through my association with John Pollock and his brother Oscar. Though I could never bring myself to believe that his brother’s inner light was switched on, John never held that against me. And I am quite sure that this attitude of mine never pained Oscar. This thesis might well carry the title “Color and Consciousness.” Joseph Tolliver taught a class by this title and I sat in on it. The cognitive impetus imparted to me in this class was what launched me on the present course of inquiry. It would seem that this impetus was so strong that it carried me way beyond anything that Joseph taught me. He is, therefore, not to be held responsible for anything written here. Had I been able to respond in a more satisfactory manner to all of Chris Maloney’s penetrating commentaries, this would be a much better dissertation than it is. Inasmuch as there are mistakes, both philosophical and linguistic, that I failed to commit in this thesis, a good part of the responsibility for that rests on Chris’s shoulders. For many years now I have had a friend and benefactor in Keith Lehrer. In directing this dissertation he assisted me in ways that go far beyond the call of duty. Had it not been for his patience, his understanding, and, above all else, his unflagging trust in me, I could not have completed this dissertation. Finally I want to thank Robert Rogers who brought me to the United States. At the time I had no very clear what I was embarking on. Now that I see things a little clearer I want to tell him how much I appreciate what he did for me.

TABLE OF CONTENTS

ABSTRACT	9
CHAPTER 1. CONSCIOUSNESS, QUALIA, AND THE FIRST-PERSON	
POINT OF VIEW	11
The Thesis	11
Clarifying the Thesis	11
Qualitative Consciousness	12
An Apology for Qualia	16
The Datum	16
Why the Datum Is Obvious	16
The Neutrality Thesis	17
Qualia and Sense-Data	20
The Qualia Quiners	21
Why the Datum Is Important	24
The Extent of Qualia	25
Why the Datum Is Baffling	27
Deflating the Bafflement	29
Background Materialism	33
Summary of the Apology	35
The Scope and Limits of this Project	36
Methodological Commitments	39
Taking Phenomenology Seriously	39
A Plea for the First-Person Point of View	44
The Restoration of Consciousness	44
Choosing the First-Person Point of View	46
The First-Person Point of View and Anti-Materialism	47
The First-Person Point of View and Anti-Scientism	49
Summary and Outlook	56
CHAPTER 2. AGAINST INTROSPECTIONISM: WHY INTROSPECTION	
DOES NOT HOLD THE KEY TO CONSCIOUSNESS	62
Explaining Consciousness through Introspection: Introspectionism	62
Connecting Consciousness And Introspection	65
The Definitional Link Between Consciousness and Introspection	65
Introspection as the Best Explanation of Consciousness	67
Consciousness and the Consciousness of Consciousness	68
The Epistemology of Consciousness and Introspection	71
Three Forms of Epistemic Access to Our Own Minds	73

TABLE OF CONTENTS—*Continued*

Making the Transition to Introspectionism.....	74
Against the Three Epistemological Theses.....	77
The Special Knowledge Thesis	78
The Argument From Materialism	79
The Argument from Conceptualization.....	80
The Argument from Deviant Reasons.....	81
Assessment of the Special Knowledge Thesis.....	82
The Fallible Knowledge Thesis	84
The Belief Thesis	85
First Case: Qualitative Beliefs without Conscious Mental States: Detached Qualitative Beliefs	88
Second Case: Conscious Mental States without Qualitative Beliefs: Detached Qualia	90
The First Appeal to Phenomenology	91
Simple Minds	98
The Connection of Introspection and Consciousness	105
Summary and Conclusion	110
 CHAPTER 3. HOW TO BUILD AN IMITATION MAN: OSCAR THE UNCONSCIOUS INTROSPECTOR	113
Pollock's Oscar as Imitation Man	114
Rational Functionalism and Consciousness.....	116
Linking Rationality and Consciousness	116
The Role of Consciousness in the Human Rational Architecture	117
How to Build Consciousness into Oscar.....	119
Fitting Oscar with Introspective Sensors	119
Fitting Oscar with Second Level Introspective Sensors.....	120
Doubting that Oscar Is Conscious	122
Perception without Sensation	123
The Lessons of Blindsight	125
Pollock's Analysis of Blindsight.....	126
A Consequence of Pollock's Analysis: No Learning without Introspection	127
The Capacities of Blindsighted Subjects	128
Blindsight and Learning: A Dilemma	130
The Argument from Blindsight	132
An Objection to the Argument from Blindsight.....	133
Reply to the Objection	133
Should we Take the Possibility of Oscar's Blindsight Seriously?	136
Two Arguments for Taking Oscar's Blindsight Seriously	137

TABLE OF CONTENTS—*Continued*

The Argument From Parsimony	137
The Argument From Mystery	138
An Objection: The Two Arguments Prove too Much	139
Two Arguments Against Taking Oscar's Blindsight Seriously	141
The Argument From Analogy	141
The (Thought)experimental Solution: Trying on Oscar	143
Conclusion	146
 CHAPTER 4. QUALIA DE-MENTED: THE CASE FOR NONMENTAL QUALIA.....	148
Giving up the Assumption that Qualia Are Mental	148
Critique of the First Appeal to Phenomenology	151
The Absence of Phenomenology	152
The Relocation of Phenomenology	154
Support for the Relocation Thesis	155
No Mental Home for Qualia	156
The Second Appeal to Phenomenology	158
The Inconclusiveness of the Appeal to Phenomenology	159
The Price of De-Mented Qualia	166
De-Mented Qualia and the Belief Thesis	168
De-Mented Qualia and Introspectionism	175
Does the Relocation Thesis Rob the Qualophile of his Subject Matter?	178
What Is It to Have De-Mented Qualia?	180
Summary	183
 CHAPTER 5. SEARCHING FOR A NONMENTAL HOME FOR QUALIA: THE CASE AGAINST NONMENTAL QUALIA.....	185
Illusions and Nonexisting Intentional Objects	185
Problems With Intentional Objects	188
Intentional Objects a la (Early) Brentano	191
Intentional Objects a la Meinong	192
Intentional Objects and Possible Worlds	196
Lewisian Realism	197
Ersatzism.....	198
The Argument from Atomism and Illusory Qualia.....	202
Direct Reductive Color Realism.....	205
The Illusion of Concrete Secondary Quality: Armstrong's Account	208
The Illusion of Concrete Secondary Quality: Some Problems.....	212

TABLE OF CONTENTS—*Continued*

On the Origin of the Illusion of Concrete Secondary Quality	214
On the Nature of the Illusion of Concrete Secondary Quality	216
First Interpretation: Total Illusion	217
Second Interpretation: Partial Illusion	219
Third Interpretation: Minimal Illusion	221
Reductive Color Realism and Illusory Colors	224
Summary	226
 CHAPTER 6. CONSCIOUSNESS: THE HAVING OF QUALIA EXPLAINED.....	229
Learning from Mistakes	229
Isolating Two Pernicious Assumptions	231
The Adverbialist Theory of Consciousness	234
The Neutral Monist theory of Consciousness	244
The Route to Self-Presentation: Identifying Act and Object	245
Applying Neutral Monism: Three Questions	248
What Is It To Have a Percept? The Bundle Theory of the Self	249
A Home for Qualia? The Bundle Theory of the Phenomenal Objects and Unborne Qualia	251
The Phenomenal Adequacy of Neutral Monism.....	257
Feeling Qualia: The Noncognitive Nature of Consciousness	258
Reid's Challenge to Noncognitivism	263
Summary	268
A Sobering Look Ahead: A Theory of Consciousness in Search of a Metaphysics	271
 REFERENCES.....	279

ABSTRACT

This is an attempt to arrive at a philosophical understanding of (qualitative) consciousness, characteristic examples of which are pains, tickles, experienced colors, sounds, tastes, and odors. Consciousness is analyzed as the having of qualia. Qualia (or phenomenal properties) are problematical because nothing (neither physical nor nonphysical, neither actual nor merely possible) can bear them. This suggests qualia eliminativism; but it is argued that qualia should be retained as properties that can be exemplified though nothing bears them. Phenomenal objects are then presented as bundles of qualia. The bundle theory of phenomenal objects is complemented with a bundle theory of the conscious subject. Qualia are crucial elements of the bundle that constitutes a conscious subject. For a subject to have a quale is for this quale to be included in the bundle that is the subject. This account makes consciousness into a noncognitive phenomenon. Having a quale is not a matter of knowing anything, believing anything, or cognizing anything in any way. It is simply to feel a certain way.

Two theses are singled out for particular critical attention. Concerning the nature of qualia, David Armstrong has argued that (color) qualia are complexes of primary qualities borne by the surfaces of (actual or possible) physical objects that we perceive or seem to perceive. More than other reductionists, Armstrong is concerned to ensure the phenomenological adequacy of his reductionist theory. This phenomenological sensibility makes his theory of qualia particularly interesting and also particularly vulnerable. Concerning the question what it takes to have qualia, introspection appears to be the chosen tool of many contemporary theorists. Using John Pollock's introspectionist account

of qualitative consciousness as a model, it is argued that introspection can play no part in an adequate explanation of qualitative consciousness.

Throughout the investigation the methodological importance of the first-person point of view is emphasized. The primary responsibility of philosophical theory of consciousness is to insure phenomenological adequacy. Straying from the first-person point of view makes it easy to forget this.

CHAPTER 1 CONSCIOUSNESS, QUALIA, AND THE FIRST-PERSON POINT OF VIEW

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 the thesis that to be conscious is to have qualia bears some explaining. The task I have set myself is to transform this slogan into a substantive philosophical thesis.

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 conceptual explication we need to have a rough preanalytic or preexplicative grip on the analysandum or

explicandum. The term consciousness is sufficiently ambiguous to require such preanalytic 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. It is unclear what the terms “quale” and “qualia” mean and many have doubted that there are any such things.

It is, therefore, necessary to begin this investigation with a rough sketch of the relevant notion of consciousness and to follow this up with 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.

QUALITATIVE CONSCIOUSNESS

Here, then, is how I propose to delimit the notion of consciousness the analysis of which will occupy the following chapters. Even if we stick to what one might call the “philosophical” uses of the term “consciousness” we find a great deal of disagreement. I shall 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.¹ Thomas Nagel has characterized this notion of consciousness most illuminatingly when he wrote:

¹ For a sampling of other philosophers who employ this notion of consciousness consult the quotes in the section “Why the Datum is Baffling,” in this chapter. The passages from McGinn and Lockwood make it clear that they employ the very notion of consciousness that I am after. Maudlin makes it clear that he too

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, p. 166]

By saying that a subject is conscious I am saying that it is like something to be that subject for that subject. And the subject's consciousness consists in that which it is like to be that subject. It is this notion of qualitative consciousness (or consciousness, for short) that I propose to analyze as the having of qualia.² We can further illuminate this notion of 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 the following mechanisms and capacities:

is interested in this sense of consciousness when he says: "Still, our best guess is that [a newly created copy of my brain] would support identical states of consciousness to mine, identical phenomenal states. In Nagel's terms, what it would be like to be the person with that brain would be just what it is like to be me. The content and structure of those subjective states of awareness are what I mean by 'consciousness'." [Maudlin, 1989, p. 408]

² Nagel's (or Farrell's, [Farrell, 1950]) notion of *knowing what it is like* has come in for a good deal of criticism. See, for example [Hofstadter and Dennett, 1981, p. 403-414], [Armstrong and Malcolm, 1984, p. 45-66], and [Lycan, 1987, p. 76-77, 142-143]. A good part of this criticism is directed at the knowledge claim that is contained in the notion of *knowing what it is like*. 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 shall not address these criticisms. But I cannot simply sidestep the following objection: "Its being like something cannot serve as a mark of consciousness for everything (also those things that are not conscious) is such that it is like something to be it. Take, for example, this marble. What is it like to be it? Well, it is to be spherical, smooth, red, etc. (This example is due to J. Christopher Maloney). Nevertheless, the marble is not conscious." How should one reply? One cannot deny that the marble is like something, for it is spherical, smooth, and red, etc. But to acknowledge that the marble is like something is not the same as to grant that it is like something for the marble to be the marble. From the "marble's point of view" things don't appear in any way at all. The marble's being spherical does not translate into its being like something to be spherical for the marble. None of the marble's properties get reflected in its being like something for the marble to have these properties. That is to say, the marble is not conscious. Thus I should 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 to be them.

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

The notion of consciousness that I am interested in is different from all of these. That which I want to understand 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. Numerous analyses of the listed mechanisms and capacities can and have been proposed that fail to shed any light on the question what it is like to be the subject possessing and exercising these mechanisms and capacities. The view that this is the appropriate way of understanding these mechanisms and capacities has gained such prominence that it has received 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, p. 309] Conscious inessentialism does not, of course, commit one to the

absurd view that there is *no* connection between possessing and exercising 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 it 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 hold true of the other proposed mechanisms and capacities listed above. They all can (and sometimes do) get exercised without its being like anything to do so.

Thus we find that conscious inessentialism derives support from two quarters. Reference to consciousness is not conceptually necessary for achieving an understanding of the proposed mechanisms and capacities. And the exercise of these mechanisms and capacities is not causally sufficient for the production of consciousness. In short, consciousness is neither conceptually nor causally essential for the possession and exercise of these mechanisms and capacities. It seems, therefore, that an investigation of consciousness should not let itself be side-tracked by an investigation of these mechanisms and capacities.

But perhaps this is going too fast. 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 and/or necessary 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 going this path. What I want to understand is not what kind of causal machinery it takes for one to achieve consciousness; 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.

AN APOLOGY FOR QUALIA

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 with 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.

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 terribly 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 focussed 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, and tastes, and bodily feelings at every moment of our waking or dreaming lives. This is a datum that stands fast for us.

The Neutrality Thesis

But 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 expressions whose purpose it is to describe the effect that 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 some 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.

But this dispute is easily avoided by simply stipulating that we are going to use the problematical terms without any metaphysical strings attached. I propose to use the terms “quale” and “qualia” (and all the other terms adumbrated 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, or whatnot. This, then, is the sense in which qualia are, for the moment, *metaphysically neutral*. In being confronted with a quale I am confronted with a property of something; but at this point we shall 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. That is, at this point I shall not draw a distinction between qualia and sensory properties of external physical objects. Thereby I take myself to have removed what is, perhaps, the main reason for rejecting the datum that stands fast for me.³

³ 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 this chapter. Price tells us that when he sees a tomato the

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 shall 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.⁴ Be that as it may, the suspicion that qualia are just resurrected sense-data is quite mistaken. The more careful discussions of sense data make it very clear that sense data are introduced as bearers of qualia. John Wisdom states the point clearly: “[S]ense data are not identical with sense-qualities but are the things which possess *or* seem to possess sense-qualities.” [Wisdom, 1963 (1934), p. 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

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, 1954 (1932), p. 3] Data of this kind Price calls “sense-data.” These sense-data are metaphysically neutral because by admitting them “[w]e 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, 1954 (1932), p. 18-19]

⁴ The horror of sense-data stems mainly from a misapprehension of their nature, viz., the belief that sense-data are nonphysical things. It is worth pointing out that such champions of sense-data as Russell and Moore denied that sense-data had to be nonphysical.

other; and should call, and have called, *only* the patch, *not* its color, size or shape, a “sense-datum.” [Moore, 1966, p. 44, fn. 2] Moore distinguishes between the sensible qualities, like a particular color, size, and shape, and the sense-datum that has those qualia, 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 the sense-datum and its sensible qualities is the crucial point of the sense-datum theorist’s defence against its most serious competitor. [Schilpp, 1968, see pp. 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 the obvious fact that we live our lives immersed in a sea of qualia.

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, p. 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, intrinsicness, 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.

But 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 this 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 both of the above worries will be settled.

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 Dennett's many puzzle cases [Dennett, 1988]. Take the dilemma of the disenchanted coffee drinker. She 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 cauliflower 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 one 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.

Can one accept the cogency of Dennett's intuition pumps and, at the same time, resist Dennett's eliminativist conclusion? I think that this is possible. For I believe that Dennett's eliminativist conclusion is not a lesson gleaned from but a philosophically loaded interpretation of the puzzle cases concerning the identity of qualia. As best I can tell, the driving force behind Dennett's reasoning is the metaphysical principle that found its expression in the slogan: "No entity without identity." It is this principle that licences the inference from the indeterminateness of questions about the identity of qualia to the conclusion that there are no qualia. But this principle is not sacrosanct. Many kinds of entities confront us with seemingly irresolvable 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.⁵

⁵ 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 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, pp. 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.

WHY THE DATUM IS IMPORTANT

“Why, if it weren’t for this ‘internal illumination’...the world would be nothing but a pile of dirt!” [Feigl, 1967, p. 138] This is how Herbert Feigl reports Einstein’s remark on the importance of the qualia. 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.

The Extent of Qualia

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. [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, p. 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 mental 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 properties 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 whole domain of what Alvin Plantinga calls the "nonsensuous phenomenology." In this area we find that qualities that go with the emotions like anger and fear; and 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.⁶ [Russell, 1978 (1921), p. 250]

⁶ 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 fascinating 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 worry about how to spell out in a psychologically credible way under which modes of presentation we present propositions to ourselves are engaging the same problem.

Once one appreciates the full scope of qualia the claim that to live a life without them would be like nothing 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 makes life worth living. This is why our original datum is so important.

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 shall indulge in the pleasure of quoting some of the most recent and most eloquent expressions of this bafflement. I do this not

only for the intellectual pleasure involved but also because 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:⁷

“[W]e 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 irresolvable enigma.” [McGinn, 1991, p. 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, p. 349]

⁷ All but one of the following expressions of the bafflement assumes that the locus of qualia is the *brain*. I have not yet committed myself to such a thesis. But the fact that the quoted authors take a more specific stance on the location of qualia than I want to take at present does not detract from the vividness with which they portray the bafflement.

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, p. 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, p.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, p. 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, p. 16]

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, p. 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 shall come back to these issues in the fifth Chapter.

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 notions. 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 this trivialization of the bafflement we avoided 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 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, p. 111]

What, exactly goes through the mind of a person who engages in such a meditation? Before presenting Lycan's analysis of what is wrong with the ruminations of one who engages in this kind of meditation I shall submit what I

take to be the correct interpretation of her thoughts. I believe that something like this goes through the mind of the baffled person. 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 red of the kind that I am currently imagining. The bafflement one experiences is similar to that which one experiences upon focussing 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 and allotropy in particular. But initially it is justified and beneficial: it is this kind of bafflement that drives us to learn more about nature. And eventually it was dispelled by a sizeable amount of scientific knowledge that enabled our understanding 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.

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, p. 111]⁸

Lycan is right in pointing out that materialism does not entail that looking at a grey brain (that has a red sensation) is like having a red quale. Materialism does not entail that; nor does any other doctrine that is not patently absurd. If the brain you watch is grey all over it will afford you a gray quale. And since a grey quale is quite unlike a red quale nobody in their right mind would expect that watching a grey brain would be anything like experiencing a red quale. A grey quale is as different as could be from a red 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, red, or whatever. And this bafflement is likely to become particularly acute when we actually look at a brain; not because the brain is grey

⁸ Lycan’s discussion of the stereoscopic fallacy does bring out one very important issue. 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? Do they qualify brain states, events, neurons or what? Or are they “in” the brain of their possessor in some completely different way? Why should it be absurd, as Lycan clearly thinks it is, to assume that if cyan qualia are in the brain the brain itself must present cyan qualia to one who looks at it with his eyes. I shall come back to this question in chapter 6.

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 qualia.⁹

Background Materialism

A 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 problematical. If, instead, one were to chose to view qualia as metaphysically prior then the existence of matter would strike one as puzzling.

If it is this background materialism that makes qualia problematical, why not simply abandon materialism? There is no short answer to the question: Why should we be materialists? 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

⁹ For more on the reasons for the supposed unintelligibility of how the brain could give rise to and/or instantiate qualia see the section "The Argument from Illusion and Illusory Qualia" in Chapter 5. This much I want to say right here: I am not defending the untenable doctrine that no two "very different" properties can be coinstantiated. But I shall hold that not 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 belong to the set of properties that cannot be instantiated together. In my argument for this last claim I shall follow Sellars's grain argument. For more on this I refer the reader to Chapter 5.

cosmological outlook. [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 it 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, pp. 53-54]

And much more recently Colin McGinn has expressed this same conviction in even stronger words:

[O]ne 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, p. 47]

McGinn too sees that his conviction that “nothing that happens in nature is inherently anomalous, God-driven, an abrogation of basic laws” is “an article of metaphysical faith.” It belongs, he tells us, with “incredulity over ghosts, telepathy, divine healing and the like.” [McGinn, 1991, p. 87]

All these sentiments I applaud wholeheartedly. We must not give up our “naturalistic sobriety” (McGinn’s term) at a 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. There are, however, two caveats I should like to register. First, I find that I loose 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.

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. I regret that I have no such account to offer. In the end I shall choose qualia over materialism.

SUMMARY OF THE APOLOGY

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—seems to me blatantly obvious. The importance of qualia resides in the fact that it is only through having them that it is like something to be you. 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. For the most part the defence of the thesis will simply consist in spelling out clearly what it is to have qualia.

THE SCOPE AND LIMITS OF THIS PROJECT

Just how narrow is this project? There are really three questions here. First, How fully do you treat the domain of qualitative consciousness? On this count the present investigation is severely limited. I shall not treat of all the kinds of qualia; not even of all the secondary qualities; not even of all the colors. For the most part I shall deal with phenomenal red only. This is a rather desperate limitation. And the hope that a similar account could be given of all qualia is, at this point, no more than a pious hope. Meekly to wave one's hand while claiming that *mutatis mutandis* the same story could be told across the board would simply be dishonest.

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 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 should 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. What is it, then, that makes the study of qualitative consciousness worthwhile? I shall give my answer to this question when dealing with the third and last interpretation of our initial question.

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. And perhaps this question does not take me to task for not addressing all of 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.¹⁰ 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 is not and might never rise to 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

¹⁰ 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, p. 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, p. 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 interesting introduction to the volume *Consciousness in Contemporary Science* [Marcel and Bisiach, 1988, pp. 1-15] and [Flanagan, 1991, pp. 307-14]. Of course the case for the scientific recognition of consciousness is not helped by the fact that some of the leading qualophiles insist on the epiphenomenal character of consciousness.

of view qualitative consciousness presents itself as an absorbing problem that forces itself upon me ineluctably. Consciousness makes it like something to be me. Without it I would be nothing to myself nor would the world be anything to me.¹¹ Echoing David Armstrong, I can say that without consciousness I would not be aware that I existed—my self would not be a self to itself. [Armstrong, 1978, see p. 67] I cannot imagine a better reason for singling out an issue for philosophical consideration.

METHODOLOGICAL COMMITMENTS

TAKING PHENOMENOLOGY SERIOUSLY

So far I have stated the thesis I want to defend; I have attempted to make the thesis tolerably clear; and I have indicated how narrow and how important I take this kind of investigation to be. Two things remain to be done to establish the thesis. First, we need to answer the question: What are qualia? (This question poses what I propose to call the *nature problem*). Second, we need to answer the question: What is it to have qualia? (This question poses what I propose to call the *having problem*). These two questions I try to answer in the following chapters; so 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.

¹¹ Enio Kaila has expressed the feeling perfectly when he wrote: "Inhalt, Sinn und Wert finden wir nur in der "subjektiven", ichzentrierten Erlebniswelt." [Kaila, 1942, p. 125]

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 are qualia? presents itself to us 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 often been suggested, nonactual physical objects. Of every proposed qualia-bearer I shall ask the following simple minded question: Can one understand what it is for an object of this kind to exemplify a quale? We shall 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 rather 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 with 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 focussed 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 a quale consists in can give us a feeling 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 the simple minded question that needs to be put to every proposed account of what the having of a quale consists in. We shall have to ask: Does this way of having a quale make it understandable how being related to a quale in this manner can be like something for the subject who has the quale in this way? We shall see that not all of the proposed having relations pass this test.

The solutions to the nature problem and the having problem are not independent of each other. The “location” of the quale will, obviously, have a strong bearing on the question what it takes for us to have the quale. 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. Finding excellent qualia-bearers that are 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. Thus solutions to the nature and the having problem will have to be evaluated as package deals, so to speak.

Since there is this interdependence between the nature problem and the having problem some remarks about how I propose to handle this interplay are called for. The discussion of the following four chapters is evenly divided between discussion of proposed solutions to the having problem and to the nature problem. Chapters 2 and 3 are focussed on one answer to the having problem. I proceed on the more or less tacit assumption that qualia are “inside.” The most

popular answer to the question of what it takes to get at inner qualia is that such qualia are had by being introspected. In Chapters 2 and 3 I argue that having qualia is not a matter of introspecting them. The focus of the discussion shifts in Chapters 4 and 5. The tacit assumption of the innerness of qualia is discarded and replaced by the explicit assumption that qualia are “outside.” I argue that the external objects do not make plausible candidates for qualia-bearers. And even if we could make sense of the idea that qualia are outside, the problem of how we come to have these qualia becomes insoluble. In the final Chapter I propose an account that is supposed to provide a solution to both the nature and the having problems in one fell swoop. The solution (if it is one) comes at a high price. It cannot be accommodated within the materialist framework.

Throughout this whole discussion I shall be guided by a certain policy. The policy is simply to be completely literally 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 thus borne. Sometimes there is a virtue to sticking to a certain pigheaded literal-mindedness in philosophical discussions. I shall not accept anything 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 discussion will show that proceeding in the proposed 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.¹² I shall follow the qualia around wherever they may take me. The neutrality thesis allows me to countenance inner as well as outer qualia. But I shall insist on the phenomenological adequacy of the resulting theory of consciousness. That is, I shall subject every proposed account of what it is to have qualia to the first-personal test of “trying it on.” I shall heed Searle’s advice: “Remember, in these discussions, always insist on the first person point of view.” [Searle, 1980a, p. 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, p. 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 stuck in its infancy. Therefore this policy is a bad one. The reply to this objection is equally well entrenched. Different subject

¹² This is one of the rare occasions where I find myself in complete agreement with Lycan’s sentiments. Defending his choice to completely neglect qualia eliminativism in a book length treatment of consciousness, Lycan has this to say: “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, Dennett, and Rey], 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, p. 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.

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.

A PLEA FOR THE FIRST-PERSON POINT OF VIEW

As the preceding section has made obvious, I undertake this investigation from the first-person point of view. The first-person point of view has come to be associated with an aura of obscurantism and mysticism. I shall, 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.

THE RESTORATION OF CONSCIOUSNESS

In prebehavioristic days of introspectionist psychology the attempt to present a theory of the mind that was not, at the same time, a theory of consciousness must have seemed like a gratuitous enterprise. The advent of behaviorism thoroughly changed the understanding of what it was to present a theory of the mind. Consciousness was demoted to a slightly disreputable side issue. But with the demise of behaviorism consciousness has made a comeback. 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 introspectionist tradition. Consciousness is no longer viewed as that which simply reveals itself in self-observation; rather it is viewed as a theoretical construct that the cognitive scientist posits in the light of certain explanatory pressures within his theoretical framework. This change of theoretical framework 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 far removed from being a monolithic enterprise. The diversity of theoretical approaches to the issue of consciousness reflects the heterogeneity of the field. The newly-discovered understanding that consciousness is a topic deserving of our attention has not lead to a corresponding agreement about what *the* problem of consciousness is. But there are at least two unifying themes in the multifariousness of voices: consciousness is understood as a *high level information processing capacity* that has to be studied from the *third-person point of view*. Neither the prevailing uncertainty about the nature of the problem nor the multiplicity of theoretical approaches can conceal the fact that much of the work in this field is informed by a shared background view. This is the view that the mind is, first and foremost, an information processing system; and that to understand the mind is to understand how it carries out its information processing tasks. When viewed from this

perspective, consciousness is most naturally understood as an information processing capacity. 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; or, negatively put, what these hypotheses share is their disregard of the first-person point of view. 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. It is clear that current theorizing about consciousness is motivated and executed in ways that differ significantly from the ones pursued in the introspectionist tradition. Whether the move from the first-person perspective to the third-person perspective can be counted as progress in our investigation of consciousness remains to be seen.¹³

¹³ My repeatedly mentioning the introspectionist psychology should not be misconstrued as an endorsement of this position. As Chapters 2 and 3 will make clear I hold that the attempt to understand consciousness in terms of introspection is mistaken.

CHOOSING THE FIRST-PERSON POINT OF VIEW

Dennett has juxtaposed the first-person and the third-person points of view with all the clarity and starkness one could wish for. [Dennett, 1987b, pp. 5-7] Which position you adopt is, according to Dennett, a matter of *choice*. For the arguments pro or con such fundamental positions as these are likely to be no more than elaborate exercises in begging the question against one's opponent. And Dennett informs us that he puts his money on the third-person point of view: "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." [Dennett, 1987b, p. 7] I propose to make the opposite choice: throughout this investigation I shall stubbornly cling to the first-person point of view.

It is not quite correct, however, to say that I want to start from a position that is directly opposed to the one from which Dennett starts. 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 chose anti-materialism and anti-science? If so he is wrong.

THE FIRST-PERSON POINT OF VIEW AND ANTI-MATERIALISM

Two of the three most vocal advocates of the first-person point of view are deeply committed to the materialist faith. For John Searle and for Colin McGinn materialism is a nonnegotiable article. And a brief look at Nagel's notorious

“What Is It Like to Be a Bat” will convince anyone that he is not the foe of materialism he is usually made out to be.¹⁴

Since counting Nagel into the ranks of the materialists may seem controversial or simply wrong, I shall briefly defend this claim. After stating his arguments against extant 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, p. 176]

That is, he takes himself to have shown that the extant versions of physicalism cannot solve the mind-body problem. But Nagel is well aware that to show this is not the same as to show that physicalism is false. He goes on to say that he believes that we actually are in possession of “some reasons to believe that sensations are physical processes....” [Nagel, 1974, p. 178] And he closes the paper with an attempt to develop a way of thinking about the subjective character of experiences that would make those experiences “better candidates for objective explanations of a more familiar sort.” [Nagel, 1974, p. 179] This is what the project of an objective phenomenology is all about: “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, p. 179] That is, what Nagel

¹⁴ In fact Nagel’s work contains numerous anti-dualist remarks. He thinks that dualism is a conceivable but implausible view. [Nagel, 1986, p. 29] More importantly, he repeatedly makes the point that the kinds of problems raised by the subjectivity of the mental prove to be no less problematical for the dualist than for the materialist. See, for example, [Nagel, 1979a, pp. 189-190] and [Nagel, 1979b, p. 201].

seems to be suggesting is that if we can replace the “faulty objective analysis of mind” by a correct objective phenomenology of the mind, then physicalism will have a chance. 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: “[C]areful 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, p. 166] Nagel’s work should be understood as a contribution towards this distant goal.

Instead of seeing “What Is It Like to Be a Bat” 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. His is a less self-assured materialism than Searle’s and McGinn’s (whose confidence in materialism seems to border on the dogmatic). But surely Nagel’s case cannot be used to argue that taking the first-person point of view seriously buys you anti-materialism. Perhaps McGinn, Nagel, and Searle are all very confused; perhaps they all should be anti-materialists, given their proclivity for the first-person point of view. But Dennett has not even attempted to show anything that strong. So I shall rest content with my appeal to authority. We have no reason to think that embracing the first-person point of view goes hand in hand with anti-materialism, for the three most prominent advocates of the

importance of the first-person point of view are more or less staunch materialists themselves.

THE FIRST-PERSON POINT OF VIEW 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 nonnaturalist 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, p. 264]

An investigation of consciousness carried out from the first-person point of view will form part of what Brentano called “descriptive psychology.” Though this label is not commonly used these days, the practice itself is well and alive in the

analytic tradition. Some of its outstanding contemporary practitioners who have written extensively on the nature of consciousness are Nagel, Searle, and McGinn, to name just a few.

Though not, strictly speaking, scientific, the results of this kind of inquiry will be a systematic, empirical, observation-based body of propositions. The empirical 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. As part of his analysis of what went wrong with the philosophy of language in the twentieth century, John Searle offers the following diagnosis: "It is part of the persistent objectivizing tendency of philosophy and science since the seventeenth century that we 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, p. 145] He then proceeds to defend the view that there is nothing unempirical about the first-person point of view and that the first-person point of view is indispensable for an adequate philosophy of language. At the heart of his defence is the claim that the rejection of the first-person point of view stems from a failure to distinguish between two senses of the term "empirical." Commenting on Donald Davidson's claim that "What an interpreter cannot on empirical grounds decide about the reference of a schemer's words cannot be an empirical feature of those words," Searle writes:

What Davidson says looks like a tautology: What can't be decided empirically isn't empirical. But the way he uses this is not as a tautology. What he means is: What can't be conclusively settled on third-person objective tests cannot be an actual feature of language as far as semantics is concerned. On one use "empirical" means: subject to objective third-person tests. On the other use it means: actual or factual. There are then two different senses of "empirical"; and the argument against the first-person case succeeds only if we assume, falsely, that what isn't conclusively testable by third-person means isn't actual. [Searle, 1987, p. 146]

Similar questions about the legitimacy of the first-person point of view arise when we turn from language to consciousness: Has the shift towards ever greater objectivity and the attendant supplantation of the first-person point of view by the third-person point of view influenced attempts to come to terms with consciousness? Theories of consciousness have not been exempt from the pressure to move towards the more objective. Intuitively it may seem that a theory of consciousness that limits itself to data that are sanctioned by the third-person point of view will have to account for a narrower range of phenomena than a theory that acknowledges both third-person and first-person data. And this intuition will lead one to question whether this third-personal bias stands in the way of a satisfactory account of consciousness. More specifically, this intuition may lead one to conjecture that the shift from the first-person to the third-person point of view transforms the scientific and philosophical investigation of consciousness in two ways. First, it results in a changed view of what one takes to be the subject matter of a theory of consciousness. Second it results in a changed view of what one takes an adequate theory of consciousness should look like. The blinkering effect of the exclusive reign of the third-person point of view has not been lost on its leading partisans. Dennett, for example, has offered the following sobering assessment:

One of the yawning chasms in the current cognitive theories of the conscious mind is the absence of any clear role for what is somewhat lamely called “affect” or “emotion,” or even just “the way things feel” or (in philosopher’s jargon) “qualia.” Concentrating as we do these days on cognition, we have tended to incorporate its all but forgotten twin, conation, the “faculty of volition and desire,” into the colorless ranks of information-bearing states; 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, p. 100]

Dennett is not, of course, guilty of simply averting 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 at all or can be accounted for from a third-personal position.

I don’t follow him in his tactical choice of sticking with the third-person view. Instead I prefer to switch viewpoints so as to see what consciousness looks like when viewed from the inside, as it were. This choice does not bespeak some form of anti-scientism on my part. First-person and third-person investigations can fruitfully coexist. As I see it, the result of the first-person investigation of consciousness will be the target of a third-person theory of consciousness. That is, a third-person, 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 problematical feature of consciousness is. Then we can attempt naturalistically to domesticate this feature by reducing it (in the broadest sense of the term) to some scientifically respectable feature of the world.¹⁵ These two

¹⁵ Something much like this view is implied by Butchvarov’s claim that “[e]ven 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, p. 264]

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-personal, naturalistic theory of consciousness. What they have argued is that we

This is also what Nagel has in mind when he says that "[n]othing is proved by the inadequacy of physicalist hypotheses that assume a faulty objective analysis of mind." [Nagel, 1974, p. 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, p. 167] Before you can have a theory of anything you need a clear notion of what your theory is supposed to be a theory of. And the discovery that a theory is false does not amount to much when it turns out that the supposed facts that the theory describes were the product of a mistaken description of the theories intended domain. Searle has his own rather harsh way of putting this matter. As far as he is concerned many of the numerous recent books that purport to be about consciousness simply aren't about consciousness at all. See [Searle, 1989, p. 193].

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 the most cursory review of his extended argument for his verdict of *ignorabimus* 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 shall 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.¹⁶

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 standpoint that is anti-science 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.

This brings my defence for the first-person point of view to an end. By adopting the first-person point of view one does not automatically subscribe to some anti-materialist, anti-scientific form of obscurantism. 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 choosing the first-person point of view makes the symbiotic coexistence at least possible. And, I suggest, that this is all one can reasonably demand of any starting point for the 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 more or less open suggestion that to make this choice is to become a mystery monger of sorts. The fact that some choose to label us with such newly coined technical terms as “the new mysterians” [Flanagan, 1991, pp. 312-314]

¹⁶ And what should we have said if it had turned out that McGinn and Nagel did argue (and argue correctly) that the first-person point of view shows that there cannot be a scientific, third-personal theory of consciousness? Should we then have shied away from the anti-scientific first-person point of view? On the contrary; we should have embraced it and bade science farewell, even if it would have made us very sad.

should not intimidate us. There is nothing anti-scientific about starting from the first-person point of view. Making this choice is not something one needs to be ashamed of in these naturalistically inclined times.

SUMMARY AND OUTLOOK

In this Chapter I have stated my thesis: To be conscious is to have qualia. And I have attempted to clarify the thesis in two ways. First, I have characterized the relevant notion of consciousness: a subject is conscious just in case it is like something to be that subject. Second, I defended an innocuous, metaphysically neutral, version of the notion of qualia. With both of these preliminary explanations in place the thesis has gained sufficient clarity to merit further discussion. I acknowledged that the thesis is a narrow one in more than one way. At the same time I insisted that the problem it addresses is of a momentous importance. These considerations naturally lead to reflections about the philosophical perspective from which I approach the problem of consciousness. In attempting to answer the question: What is it to have qualia ? I am going to pursue a certain strategy. The crucial test for any account of consciousness is its capacity to save the phenomena. Therefore the objective of phenomenological adequacy will guide all my critical and constructive steps on the way towards an account of consciousness. As this insistence on phenomenological adequacy indicates, the investigation is undertaken from the first-person point of view. Thus it seemed fitting to close this Chapter with a defence of the first-person point of view against some undeserved criticisms.

In Chapter 2 I ask the following question: What makes the belief that consciousness can be explained by appeal to introspection so attractive? I argue

that the primary reason for the attractiveness of this mistaken belief (introspectionism) stems from a confusion between consciousness and our knowledge of it. I argue that this confusion is nourished by a variety of theses about the special kind of epistemic access we have to our conscious mental states. Even the weakest of these theses, the belief thesis, cannot be upheld. The thesis that one believes oneself to be in a particular qualitative mental states just in case one is in this qualitative mental state fails in both directions. Having a qualitative beliefs is neither sufficient nor necessary for one's being in the corresponding conscious state. With the downfall of the belief thesis introspectionism loses much of its intuitive appeal. The link between consciousness and introspection is further weakened once one notices that consciousness has no proprietary phenomenology of its own. Introspection neither explains consciousness nor does it add to our consciousness. The two phenomena have nothing to do with each other. We should resist the seductive power of introspectionism.

In Chapter 3 I continue to pursue the case against introspectionism by looking at a concrete instance of an introspectionist theory of consciousness. Not all introspectionists are confused introspectionists. Pollock's version of the thesis is immune to the kinds of criticisms considered above. Pollock presents us with a detailed model of our cognitive architecture and suggests that qualia can easily be accommodated within the proposed structure of sensors and introspective sensors of sensors. In this Chapter I try to show that we have good reasons to resist Pollock's proposal to identify qualia with certain features of his framework of hierarchical sensors. I suggest that Pollock's theory of consciousness does not pass the "trying on" test. I shall try to persuade you that being the type of being that Pollock describes is like nothing.

Chapters 2 and 3 are primarily addressed to the having problem. In these chapters I try to show that the favorite way of solving the having problem, viz. by appeal to introspection, is a dead end. The discussion in these two chapters proceeded on the unspoken assumption that qualia are inner properties; properties that are instantiated at a site to which inner perception has access. Chapter 4 challenges this tacit assumption by introducing the thesis that qualia are neither inner nor mental but, instead, respectable properties of respectable external objects. Given the neutrality thesis about qualia this possibility has to be taken seriously. Thus in Chapter 4 our focus shifts away from the having problem and towards the nature problem. Much to my chagrin I find myself forced to acknowledge that an appeal to phenomenology cannot rule out this relocation of qualia. In the second half of Chapter 4 I try to assess the consequences of de-mented qualia (that is what you get when you move qualia out of the mind and onto the external nonmental objects). Some of these consequences support my case. De-mented qualia afford further ammunition against the belief thesis and against introspectionism. But at the end of the Chapter 4 I have to admit that I have no idea what it would take to have de-mented qualia in a way that would make it like something to have them. That is, if qualia are features of external nonmental objects then I do not know how consciousness is to be accounted for. This provides me with a good reason for a critical look at the thesis of de-mented qualia.

Chapter 5 argues that the thesis of de-mented qualia is false because the nonmental world does not contain any objects that could reasonably be pressed into service as qualia-bearers. I defend the view that neither actual nor possible physical objects can exemplify qualia in a manner that would allow us to have

them in the right kind of way. Some of the proposed qualia-bearers are such that it remains completely obscure what it would be for an object of this kind to exemplify a quale. For other proposed qualia bearers this problem does not arise, but it remains unintelligible how we could possibly experience the qualia they exemplify. The difficulties that arise when one attempts to relocate qualia have, of course, not gone unnoticed by the proponents of this move. A popular maneuver executed to extract oneself from these kinds of difficulties is to declare qualia (at least partly) illusory. I lavish particular attention on this tactic and hope to show that it does not help in finding a home for de-mented qualia in the nonmental world. The friends of de-mented qualia had many reasons to relocate the qualia. One of them was that the “inside,” the mind, does not contain any qualia-bearing objects. I go along with this assumption. But now I have also argued that the nonmental world does not contain any qualia-bearers. So if there are no qualia bearers and qualia need to be borne, then there are no qualia. But since this conclusion is unacceptable one of the assumptions that leads up to it must be given up. That is what I do in the next Chapter.

In Chapters 2-5 we have learned what qualia aren't and what it isn't to have them. That may not seem like much. But it does help to point us in the right direction. We now know what pitfalls to avoid if we are going to have an account of what it is to have qualia. In Chapter 6 I present an account of consciousness, i.e., of the having of qualia, that is inspired by the desire to avoid the mistakes that have vitiated the other accounts that were discussed. What is needed are unrellocated, unborne qualia that are had, not by introspecting them, but by their presenting themselves to us. Such an account of consciousness can be had if we are willing to give up two assumptions: First, the assumption that

qualia need objects or substances to inhere in. Second, the assumption that the having of a quale is a matter of our relating ourselves to the quale through an act. The adverbialist theory of consciousness can be understood as a theory that does give up both of these assumptions. It is an attractive theory. But I shall reject it, in the end, as not being able to give a phenomenologically adequate account of consciousness. In its stead I propose what I term the neutral monist theory of consciousness. I propose that qualia are genuine properties, but properties that need no bearer to be instantiated. The phenomenal object is understood as a bundle of such unborne phenomenal properties. The bundle account of the phenomenal object is complemented with a bundle account of the conscious subject. For a subject to have a quale is for this quale to be part of the bundle that is the subject. This is the sense in which qualia are had by means of self-presentation. Membership in an appropriate bundle is all it takes for the quale to be had. No act of any kind is required on the subject's side. I argue that the neutral monist theory of consciousness is preferable to the adverbialist theory. Both theories avoid the pitfalls that we uncovered in the previous chapters: both theories avoid the positing of problematical qualia-bearers and the commitment to problematical acts that let us have our qualia. But the neutral monist account is at an advantage in that it gives a phenomenologically more credible account of what it is like to have qualia. One interesting consequence of the neutral monist theory is that having qualia, or being conscious, is a noncognitive state. That is, being conscious is not a matter of knowing or believing anything. To my mind this is another welcome consequence of the neutral monist theory of consciousness. I end the Chapter and the dissertation on a less upbeat note. I cannot see how a materialistic metaphysics can

accommodate the neutral monist theory of consciousness. I regard this as a major liability of the theory. But if I have to choose between giving up the only satisfying account of consciousness that I know of or giving up an otherwise attractive metaphysical framework that lacks the resources to account for consciousness in a satisfactory manner, I choose to give up the metaphysics. I end by proposing, in a very tentative and sketchy manner, an outline of a metaphysics that can accommodate the neutral monist theory of consciousness. It will come as no surprise that this metaphysics is Neutral Monism.

CHAPTER 2

AGAINST INTROSPECTIONISM: WHY INTROSPECTION DOES NOT HOLD THE KEY TO CONSCIOUSNESS

EXPLAINING CONSCIOUSNESS THROUGH INTROSPECTION: INTROSPECTIONISM

Much current thought about consciousness is guided by the conviction that introspection holds the key 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 shall use the label *introspectionism* to refer to this kind of view that holds that consciousness can thus be explained in terms of introspection. I shall argue that the connection between consciousness and introspection is much more tenuous than introspectionism supposes. More particularly, I shall maintain that consciousness (understood as 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.

I see one possible (and problematical) connection between introspection and consciousness. It is this: if there is such a thing as introspection, then we might expect this quasi-perceptual process to be attended by a phenomenology of its own. Whether such a phenomenology of introspection exists is questionable. In this respect the phenomenology of consciousness is quite unlike the rest of our qualitative consciousness and more problematical than it. But even if we assume

that introspection comes with a phenomenology of its own, we still have no reason to believe that the addition of this peculiarly introspective dimension of consciousness will throw any light on the puzzles surrounding our more robust nonintrospective phenomenology that is associated with our other sensory modalities. To discover that introspection has its own proprietary qualia is not to solve the problem of qualitative consciousness but to discover that this problem is even more extensive than anticipated. As far as I can tell, there is no such proprietary phenomenology of consciousness; and even if there were one, that would not advance the cause of introspectionism.

The cognitive revolution freed contemporary theorizing in the philosophy of mind from the behaviorist phobia of inner processes and mechanisms. Questions about the nature and function of introspection acquired a new respectability. And nothing stood in the way of explaining introspection by appeal to inner mechanisms and processes. One could, again, attempt to understand introspection as perception taken inside. The natural way to go was to posit a second, inner, perceptual system. On this view consciousness turns out to be “inner perception” in the most literal sense of these terms. The construct most frequently employed by materialist and functionalist theoreticians of introspection was and is that of the inner eye, the self scanning or self-monitoring device, the higher level sensor. Just as our outer perceptual system informs us about the states of our outer environment, our inner perceptual system informs us about the states of our “inner environment.” We thus end up with a hierarchy of sensors (or detectors, or “eyes”) in which the higher level sensors monitor (among other things) the functioning of the lower level sensors. On this type of account our capacity to introspect our mental states is no more

puzzling than our capacity to perceive our surroundings. While it may prove difficult to work out the details of the implementation of such a hierarchy of sensors in our nervous system, the underlying idea is straightforward and attractive. In any case, I shall not here try to criticize the hierarchical model of introspection.¹⁷ I want to focus on the question whether this kind of machinery can be used to throw light on the nature of consciousness. Theories that attempt to explain consciousness by appeal to the hierarchical model of introspection I call *hierarchical theories of consciousness*.¹⁸ What I shall argue, then, is that hierarchical theories of consciousness are misconceived.

Whether the failure of introspectionism reflects badly on the hierarchical model of introspection is a separate issue. One's view of whether and to what degree the breakdown of the hierarchical theory of consciousness weakens the hierarchical model of introspection will depend on one's view on the nature of the connection between these two theories. Since I hold that introspection and consciousness are not relevantly connected, I believe that the hierarchical theory of introspection survives the collapse of the hierarchical theory of consciousness unscathed.

But the issues of whether and how introspection and consciousness are linked to each other are controversial and therefore deserve a fuller discussion.

¹⁷ 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 theories of introspection. See also [Lycan, 1987, chapter 6] and, especially, [Hill, 1991] for a defence of the "volume control" model of introspection. For a critique of the hierarchical model of consciousness see [Lyons, 1986, chapters 3&4].

¹⁸ Hierarchical theories of consciousness constitute only *one* form of the more general thesis of introspectionism. Theories of consciousness that relied on the volume control model would, for example, constitute another form of introspectionism. Since I shall only be concerned with hierarchical theories of consciousness, I shall allow myself to use the terms *introspectionism* and *hierarchical theory of consciousness* interchangeably.

CONNECTING CONSCIOUSNESS AND INTROSPECTION

We have noted that numerous theorists take the notions of consciousness and introspection to be closely associated; but why this should be so we have not seen so far. What, then, are the reasons for supposing that introspection (more particularly, the hierarchical model of 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. But I shall argue that these three are insufficient to support the view that the issues of consciousness and introspection are inseparable.

THE DEFINITIONAL LINK BETWEEN CONSCIOUSNESS AND INTROSPECTION

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. On the Lockean view that “consciousness is the

perception of what passes in a man's own mind," [Locke, 1982 (1689), II,I,19] the connection between consciousness and introspection is simply a matter of definition. The Lockean view has inspired numerous philosophers to explore the connection between consciousness and perception, more precisely, the connection between consciousness and inner perception or introspection. That is not to say that there exists a large number of unreconstructed Lockean about consciousness. But we should not underestimate Locke's influence, even if few contemporaries would go so far to unhesitatingly endorse Locke's simple definitional link between consciousness and introspection. By introducing the term "consciousness" in the way he did, Locke started the tradition of treating consciousness and introspection as closely related topics (no matter how that relation was spelled out by the individual philosopher in the tradition).

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 definitional equation of consciousness with 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 proposed definition. The definitional equation of consciousness with introspection may be intended to achieve one or the other of two very different goals. It might be put forward as a stipulation; the 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 shall argue later on, the particular notion of consciousness that we are interested in—qualitative consciousness—is not captured by equating consciousness with introspection.

INTROSPECTION AS THE BEST EXPLANATION OF CONSCIOUSNESS

But there are also less direct ways of bringing together consciousness and introspection. If one takes consciousness to consist in, for example, the capacity to integrate and process information from various information sources and to use this information in guiding one’s behavior, then one will naturally assume that an explanation of the introspective access we have to our various perceptual and cognitive processes 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 is 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 properties 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.

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. The third consideration that may tempt one to look towards introspection in the hope of finding a solution for problems about consciousness is rooted in a confusion of the thing known

(consciousness) with the knowing of the thing (the consciousness¹⁹ 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, p. 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 nonetheless logically distinct from the other states. This distinction is crucial to the understanding of consciousness... . [Lloyd, 1989, p. 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 a rock with our cognition of it (our belief about it, or our knowledge of it) matters stand differently when what we want to investigate are states that are, in some sense, of the same kind as the states by means of which we gain access to them.

But just pointing out that our conscious states as well as our cognitions of them are mental does not go deep enough as an explanation of the propensity that many of us have to substitute discussions of introspection for discussions of consciousness. The temptation to change the topic springs from a correct insight. What is mistaken is merely the lesson that many have drawn from this insight.

¹⁹ Note that in this use the “consciousness” does not refer to qualitative consciousness but to some kind of cognitive internal access to qualitative consciousness.

The correct insight consists in this: unlike rocks, tables, chairs, and most other subject matters, *consciousness has built into it a characteristic relation to an apprehending mind*. The mistake consists in misinterpreting this directedness toward an apprehending mind in a manner that yokes consciousness to introspection.

The phrase “consciousness has built into it a characteristic relation to 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 this cryptic phrase. For I shall argue that the characteristic relation of consciousness to 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 shall not, of course, deny that we are conscious, i.e., that we have qualia. But I shall maintain that this having of qualia is not a matter or one’s knowing or believing anything.

To spell out what I do (and what I don’t) mean by this phrase is, in a way, the main objective of the following chapters. It is therefore futile to attempt to say, right here and right now, in a sentence or two, what I am going to say, in so many more words, in later chapters. I can, however, provide a hint of what I am trying to express by saying that consciousness has built into it a characteristic relation to 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 presentational mental activity on the part of the conscious subject. Since we have qualia and 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. In coming to be presented with a quale I have not exercised any mental activity, introspective or otherwise, and I do not come to know or believe anything. The upshot of all this is that instead of saying that consciousness has built into it a characteristic relation to an apprehending mind it would be much less misleading to say something like this: *Consciousness makes itself felt to the conscious person*. But further elucidation of this claim will have to wait for later.

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. In the remainder of this chapter I shall 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 to an apprehending mind*.

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 express so succinctly. In

what follows I want to elaborate this claim. I shall do this by considering three epistemologically motivated theses about the nature of consciousness. In the minds of those who are tempted to endorse theses such as these there is a very short route leading from these epistemologically motivated theses to an introspectionist explanation of consciousness. After stating the three epistemologically motivated theses about consciousness I shall lay out, as sympathetically as I can, the reasoning that takes you from there to introspectionism.

The critique of this epistemologically motivated route to introspectionism divides in two parts. First we can ask whether the passage from the epistemologically motivated theses about consciousness to introspectionism is really as safe as it is made out to be. Second, we can ask whether there are any good reasons to accept the three epistemological theses about consciousness. I shall argue that the epistemologically motivated push towards introspectionism fails on both counts.

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. This first step consists in showing the three epistemological theses to be misguided interpretations of the characteristic relation of consciousness to an apprehending mind that are unable to support an introspectionist theory of consciousness. The second step will have to consist in a positive account (that does not involve introspection) of what this characteristic relation to an apprehending mind consists in. I shall now attempt to take the first, negative, step. Step two, providing a positive account of the directedness of consciousness towards an apprehending subject, I shall attempt in Chapter 6.

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 knowableness. 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 theses—the three *epistemological theses*, for short—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 epistemological 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.

The special knowledge thesis, the fallible knowledge thesis, and the belief thesis are three ways in which one might try to interpret the characteristic relationship that consciousness bears to an apprehending mind. According to these three theses being conscious of a qualitative mental state is, respectively, a matter of knowing the state in a special manner, of knowing it fallibly, or of invariably having a belief about the qualitative state. Each one of the epistemological theses specifies (at least) a necessary condition for what it takes for one of our mental states to be 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.

MAKING THE TRANSITION TO INTROSPECTIONISM

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? The following answer will readily present itself to one who subscribes to the special epistemic status of qualitative mental states: 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 being conscious of 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.

This, then, is the reason for our interest in the 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 inner sense. In this manner each one of the three theses that were offered as possible interpretations of the directedness of consciousness toward an apprehending mind 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. The opponent of introspectionism will therefore have to find a fault with the reasoning that takes one from the epistemic theses about consciousness to the conclusion that introspection is at least a necessary condition of consciousness. Two critical routs present themselves: one can attack the validity of the move from the epistemological theses to the

introspectionist conclusion; and one can attack the truth of the epistemological theses from which the argument proceeds. I shall pursue both lines of criticism, starting with the first.

SEVERING THE CONNECTION

This manner of accomplishing the transition from the three epistemically driven theses about consciousness to introspectionism relies upon a substantive and problematical assumption. The problematical assumption—call it the *access assumption*—is this: 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. If the access assumption is granted the thesis that we are cognizant of every qualitative state of ours does entail that every qualitative state of ours is introspected. I feel that one who is inclined to ascribe a special epistemic status to qualitative mental states is prone to make the access assumption. Thus I do not expect my chosen opponents to object to my way of characterizing the passage from the three epistemological theses about consciousness to introspectionism.

This does not mean, however, that we have to accept the access assumption. As the discussion of the special knowledge thesis below will make clear, I reject the problematical assumption. There I endorse the view that it is possible to come to hold beliefs about one's qualitative states without basing these beliefs upon one's introspectively accessing the relevant qualitative states. If this is indeed a possibility (and this issue is controversial) then the passage from the three epistemological theses about consciousness to introspectionism is no longer guaranteed by an entailment relation.

While this consideration deals a hefty blow to our opponent, it does not manage to completely dispel the argument's persuasive force. The move from the epistemological theses to introspectionism has a certain plausibility, even if the entailment fails. Could not our opponent attempt to motivate introspectionism by appeal to the epistemological theses without relying on the access assumption? I think that this is possible. Giving up the access assumption blocks the entailment, but it does not sever all ties between the three epistemic theses about consciousness 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 view that being conscious is a matter of our cognizing our qualitative states in a special way strongly suggests the mistaken introspectionist account of consciousness. According to this weakened argument introspectionism is not mandated but it does, nevertheless, appear as an eminently plausible approach to the problem of consciousness. This is more than we want to grant the introspectionist.

A G A I N S T T H E T H R E E E P I S T E M O L O G I C A L T H E S E S

The first attack on the argument that takes us from the three epistemological theses to introspectionism is not decisive. Our opponent may want to hold fast to the problematical access assumption. And even if he is willing to give up the problematical assumption, he can restate his argument in a weakened form that is

no longer dependent upon the contested access assumption. In order to avoid unnecessary controversy in what follows I shall not assume that our first attack on introspectionism was successful. To further strengthen the case against introspectionism we shall, therefore, have to turn to a scrutiny of the premises from which the argument for introspectionism start, i.e., to the three epistemologically motivated theses about consciousness. If the epistemological theses are wrong, the introspectionist will loose, no matter how the debate about the entailment relation turns out.

THE SPECIAL KNOWLEDGE THESIS

Before actually launching the discussion of the special knowledge thesis I need to say a word in defence of the order in which 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 should say this: it is interesting to see the special knowledge thesis fail on its own turf, 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, p. 2]

The tradition of making extravagant claims about 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 numerous: 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, among others, D.M. Armstrong, Keith Lehrer, and John Pollock.

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, p. 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 rather formidable challenge to the claim that our conscious mental states have a privileged epistemic status.

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, p. 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 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.²⁰

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, were one to hold them 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-

²⁰ See Lehrer’s discussion in [Lehrer, 1990, pp. 67-68] Lehrer does acknowledge that there exists a very small number of propositions of which we have special knowledge; see *op.cit.* pp. 47-48.

introspective) grounds. Therefore these beliefs lose their claim to a 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, p. 54]

Pollock's point is 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 a 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, p. 61]

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

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.²¹ Accordingly, these arguments show that the directedness of consciousness towards an apprehending mind 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 characteristic relation to a mind that characterizes our conscious mental states.

This result also 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 mental states in a special way and knowing them in this way constitutes our being conscious of them. We cannot know our mental states in this 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 *special* 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 have now to turn.

²¹ 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 stem from somewhat different considerations.

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 shall now argue that the *fallible knowledge thesis* is indefensible as well. This time, I shall base this 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 the case of the special knowledge, the fallible knowledge case loses nothing of special interest if it falls together with the belief thesis. Let me back this claim up 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, 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 externalism and internalism yield the result that the fallible knowledge thesis and the belief thesis collapse into each other equivalent, given certain plausible assumptions.

Since we are interested in these two theses only insofar as they serve as a vehicle for the advancement of introspectionism, I shall not pursue the question of their epistemological relationship any further. For our purposes we may rest assured that we shall miss nothing of importance by skipping the fallible knowledge thesis and turning directly to the belief thesis.

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 our conscious mental states are invariably accompanied by a belief in their existence 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 we take the thesis precisely to assert. 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 belief²² 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 with each other (simple identity thesis) or as partly constituting each other (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.²³ And, finally, one might want to understand the belief thesis as the functional identity thesis²⁴: the thesis that the functional identity of a conscious state type is partially constituted by the fact

²² 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, p. 189]

²³ The terms "identity thesis" and "containment thesis" are Pollock's. See [Pollock, 1986, pp. 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.

²⁴ See, for example, [Shoemaker, 1975, p. 189]

that it typically gives rise to a belief that one presently is in a conscious state of this particular type.²⁵

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 having the relevant qualitative belief is not a necessary condition upon our having the corresponding conscious state. We need to show that it is false that no mental state of ours can be conscious unless we have a belief about it. In other words, we shall have to show that we can and do have qualitative mental states that are unaccompanied by the correlative qualitative beliefs. I shall 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 task is, therefore, to defend the *thesis of detached qualia*. But before taking up the case

²⁵ I want to comment briefly on the epistemic relations that obtain these various theses. 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.

for detached qualia I want to briefly investigate the *thesis of detached qualitative beliefs*. I shall start with the thesis of detached qualitative beliefs because it lets us see that what may well be the most natural construal of the belief thesis is untenable. 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 states a necessary condition on our having a conscious mental state.

FIRST CASE: QUALITATIVE BELIEFS WITHOUT CONSCIOUS MENTAL STATES: 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 that we are currently having a particular qualitative state when this is false. The argument from deviant reasons 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 that the intuition that qualitative beliefs might occur unaccompanied by the corresponding qualitative states is robust. 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. 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. It 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 them to exhibit. 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: "is that special inner light really turned on, or is there nothing but darkness inside?" [Dennett, 1987, p. 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. 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 the belief 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. And to present one of one's own mental states to oneself is just to introspect this mental state. This is 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 defence of the thesis of detached qualia.

SECOND CASE: CONSCIOUS MENTAL STATES WITHOUT QUALITATIVE BELIEFS: DETACHED QUALIA

In this section I shall consider two reasons for the belief that conscious mental states can and do occur without a corresponding qualitative belief. First, I shall maintain that a phenomenological examination of the nature of our experience shows the existence of qualitative mental states, unaccompanied by a corresponding belief. Second, I shall 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.²⁶

THE FIRST APPEAL TO PHENOMENOLOGY

The first thing I should 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.

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 is, at best, 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

²⁶ From now on I shall 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.

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. Because 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 shall cite in my attempt to buttress the phenomenological finding reported above come to this issue from varying angles and they report their findings in terms that differ from those that I have chosen. Nevertheless it will become obvious that there is a unifying theme running through their reports. All of them, or so I shall maintain, draw the distinction between qualitative states on 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 shall enlist them all to support the thesis that I have dubbed the thesis of detached qualia.

Witness 1. William Alston comes to the issue via a discussion of perception. In the course of developing his theory of perception, the Theory of Appearing, William 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, p. 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, p. 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, p. 37] As Alston sees it, this nonconceptualized noncognitive 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, p. 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.²⁷

Witness 2. John Pollock comes to the issue from an epistemological point of view. 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, p. 61] He elaborates on 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:

²⁷ Compare Fred Dretske's analysis of simple seeing. There is, he maintains, a way of seeing that is “logically independent of whatever beliefs we may possess.” [Dretske, 1969, p. 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, p. 9]

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). Applying this to our ability to turn our attention inward and focus on our sensory experience, what that amounts to is the ability to introspectively sense our sensory experience, but that does not automatically involve our having beliefs about that sensory experience. Psychologically, our perceptual beliefs about the objects in our immediate environment are caused by our having sensory experiences of certain types, and by introspection we can readily become aware of having sensory experiences, but that does not mean that we also have the beliefs that the sensory experiences are *of* the requisite types. Very often we are not even aware what the requisite types are. [Pollock, 1986, p. 62-63]

I do not want to go along with everything that is stated 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 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. And the fact that Pollock makes the thesis of detached

qualia into the cornerstone of his critique of foundationalism makes it evident that he regards it as a thesis of great philosophical importance.

Witness 3. In his book *Thought and Experience* Peter Hess has made the crucial distinction between “the cognitive and noncognitive aspects of the mental” [Hess, 1988, p. 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 phenomenological sensitivity to the distinction between qualitative states and qualitative beliefs. A number of positions that I shall 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 noncognitive 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, p. 34] The following passage elaborates the distinction a little more fully and also introduces the term “self-presentation” that I shall make much of in later chapters:

[I]t 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, p. 126]

Thus we have found another philosopher who defends the thesis of detached qualia on the basis of phenomenological insight. Witnesses could be multiplied;²⁸ but I shall resist the temptation.²⁹

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. As I have presented the matter, introspectionism is not grounded in a phenomenological finding. 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.

²⁸ For further references see [Moser, 1989, p. 80-88 (especially footnote 26 on pp. 86-87)]

²⁹ Those readers who feel that neither my appeal to phenomenology nor my appeal to authority has succeeded in establishing the purportedly robust phenomenological fact of detached qualia are referred to Chapter 4 where the purported phenomenological findings will be radically reevaluated in the light of new and conflicting arguments.

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 of our qualitative states.

By attacking the belief thesis on phenomenological grounds I have attempted to block this route to introspectionism. Since there may be other routes to introspectionism, 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 upon him to show on what other grounds his thesis recommends itself to us.

SIMPLE MINDS

Up to this point the argument against the thesis of introspectionism proceeded in an indirect manner. I argued that the acceptance of the belief thesis (or of one of the stronger variants of this thesis) is what makes the connection of the issues of

consciousness and introspection plausible. Then I tried to show that the belief thesis (and *a fortiori* its stronger relatives) does not stand up to scrutiny. And I concluded that this shows the basic idea of introspectionism to be quite unsupported and that it does not merit our 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 or the hierarchical theory of consciousness. 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 (that is a direct attack on introspectionism) is harder to defend than the weaker version (that attacks introspectionism via a critique of the belief thesis).

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

If the hierarchical theory of consciousness 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 “The Messes Animals Make in Metaphysics”³⁰

³⁰ The paper by this wonderful title is cited by [Radner and Radner, 1989, p. 7].

we may be lead 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.³¹

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 inner eye, no self-scanning device, no sensor-sensors. If animals lack all means for inner perception and thus lack the capacity to introspect, then the hierarchical theory of consciousness predicts that animals are unconscious.³² Therefore the existence of animal consciousness falsifies introspectionism understood as the hierarchical theory of consciousness.

³¹ For an exposition (but not an endorsement) of this kind of argument see [Carruthers, 1989].

³² As pointed out above, this discussion is focussed 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 argument from simple minds.

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 had would be 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 kind of 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 shall, therefore, focus on the strong version of the argument from simple minds.³³

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

³³ 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.

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 lead to doubt that animals do enjoy qualitative experiences. Here is how Lloyd puts this point:

[S]imple 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, p. 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.³⁴

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,

³⁴ There exist other more involved arguments in support of this particular form of skepticism. Peter Carruthers, for example, has recently argued the case against animal consciousness. [Carruthers, 1989] 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, p. 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 shall not pursue it here.

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, p. 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, p. 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 existed detached qualia, 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. [Griffin, 1984, pp. 186-191] If Griffin is right I see no way of upholding the argument from animal consciousness. The set of animals to which I should 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 detached qualia dwindles to nothing.

Of course we could also chose to view Griffin's incontinent proliferation of self-reflective animals as a reductio of the thesis of natural psychologists. 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 "[m]ost men seem incapable of acquiring it in any considerable degree." [Reid, 1785 (no year), p. 42] But surely Griffin is off by just as much when he praises the introspective powers of the ant.

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. 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?"³⁵

THE CONNECTION OF INTROSPECTION AND CONSCIOUSNESS

In this Chapter I have attacked the connection between consciousness and introspection as envisioned by introspectionism or, more precisely, as envisioned by the hierarchical theory of consciousness. I have tried to show that introspection does not guarantee consciousness. And, more importantly, I have attempted to free consciousness from its dependence on introspection. Accordingly introspection is neither necessary nor sufficient for consciousness. But despite this insistence on the independence of these two phenomena we should expect the following connection between consciousness and introspection. The hierarchical theory of consciousness construes introspection as a perceptual faculty. And *qua* perceptual faculty introspection should come with a phenomenology of its own. That is, just like vision and hearing come with a characteristic phenomenology of their own, we should expect our introspections to be accompanied with characteristically introspective qualia. This is definitely not the kind of nexus between qualitative consciousness and introspection that the hierarchical theory of consciousness postulated. But it is a link, nevertheless, and therefore invites investigation. Moreover, the question of the phenomenology of introspection has the following (indirect) bearing on the hierarchical theory of

³⁵ Chuang-Tze's story is related by Russell in [Russell, 1954 (1927), p. 180].

consciousness. The hierarchical theory of consciousness relies on the perceptual model of introspection. A blow to the perceptual model of introspection would, therefore, also be a blow to the hierarchical theory of consciousness. I shall now argue that we can deal such a blow to the perceptual model of introspection. Introspection does not have the requisite phenomenology. Therefore introspection should not be construed as a perceptual process. And the hierarchical theory of consciousness is weakened inasmuch as it relies on the perceptual model of introspection.

The hierarchical theory of consciousness presents consciousness and introspection as intimately related. Without introspection there is no qualitative feel, no phenomenology. As Armstrong—one of the most outspoken defenders of the hierarchical theory—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, p. 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 (that are not introspectings). The pivotal role that introspection plays in the production of our visual, auditory, tactile, etc., phenomenology (according to the hierarchical

theory of consciousness) 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 that the two issues of the phenomenology of introspection and the phenomenology that introspection produces have often been run into one.³⁶ Since I have already argued that introspection cannot create our (nonintrospective) phenomenology, I shall now be concerned exclusively with the question of the phenomenology of introspection.

The account of qualitative consciousness that I shall present in Chapter 6 presents a very different picture of the connection of consciousness and introspection. On that 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. On this view the existence and the feeling of a quale are the same and “both” are prior 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 for the production of any of the visual, auditory, olfactory, gustatory, or tactual qualia nor, it would seem, are there any other kinds of qualia left over to guild and stain our inner perceptions. In introspecting a red qualitative state the phenomenal red belongs to the introspected qualitative state and not to the

³⁶ There are notable exceptions. Pollock, who has presented the most detailed hierarchical theory of consciousness to date, has explicitly denied that introspection as such comes with a qualitative dimension of its own: “There are limits to introspection, and in particular the process of second-order introspection itself cannot be introspected. (Human beings have no third-order introspective sensors). This corresponds to the fact that unlike visual perception, introspection does not have a qualitative feel.” [Pollock, 1989, p. 14]

introspecting of the qualitative state. And introspecting the red qualitative state does not add any further qualia to the qualia that go with the qualitative state itself, red or otherwise.

To remove qualitative phenomena from the purview of introspection is not to belittle the role of introspection. Introspection allows us to *cognize* our noncognitive qualitative states. But it does not do that in a quasi-perceptual manner, i.e., it does not do that by producing percepts of our percepts or feelings of our feelings. Sidney Shoemaker has argued this position very forcefully:

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. There is something it is like to be in pain. And because being in pain and feeling pain are one and the same thing, there is something it is like to feel pain. If one holds the act-object conception of sensation, one may be tempted to equate the introspective awareness of the pain with the feeling of the pain: 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, whether or not one adopts the act-object conception. Feeling pain and being in pain are, to repeat, the same thing; and the introspective knowledge that I am in pain is at the same time the introspective knowledge that I feel pain. And it is certainly not the case that I feel my self feeling pain; there is not a feeling of the feeling that is something over and above the feeling of pain. 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, p. 114]

To deny the existence of a phenomenology of introspection³⁷ is not, of course, to deny the existence of introspection. But, given the uncontroversial assumption that “if introspection were really a form of sense-perception we should expect it to bring its own distinctive phenomenology” [McGinn, 1982, p. 51], the lack of a distinctive phenomenology of introspection must count against the perceptual model of introspection.³⁸ On these grounds I am led to accept Searle’s claim that “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, 1989, p. 123]

I conclude that there exists no relevant connection between one’s possessing qualitative consciousness and one’s exercising or having the ability to introspect. Introspection is not needed to make us have or feel our qualia; nor does it have a phenomenology of its own. This is the primary result that I have tried to establish. A second conclusion one can draw is this: not having a phenomenology of its own, introspection should not be construed as a perceptual process. Therefore the hierarchical theory of consciousness is further weakened inasmuch as it presupposes this construal of introspection.

Finally I want to forestall one possible misunderstanding of this critique of introspection. I have argued that introspection has no role to play in the

³⁷ For similar denials of a *sui generis* phenomenology of introspection see, for example, [Lyons, 1986, p. 96]: “Any experiential qualities in introspecting a patch of blue seem to be borrowed from the first-level experience, which is perceiving the blue;” and [McGinn, 1982, p. 51]: “The only phenomenology attending such awareness [i.e., an introspective awareness] comes from the *object* of the awareness.”

³⁸ Other authors who explicitly make this assumption include [Lyons, 1986, p. 96] and [Shoemaker, 1986, p. 114]. It should be pointed out, however, that this assumption has not gone unchallenged. The phenomenon of blindsight has led some authors to countenance the notion of “perception *sans* sensation.” [Humphrey, 1986, p. 57]. For more on this issue see the discussion of the phenomenon of blindsight in Chapter 3.

explanation of qualitative consciousness. But in pointing out these limits of introspection I have not been trying to insinuate that introspection has no role to play in our cognitive economy as a whole. A nonperceptual faculty of introspection will have an important role to play. 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.

SUMMARY AND CONCLUSION

In this Chapter I have attempted to show that there exist no interesting connections between consciousness and introspection. Neither can consciousness be explained by an appeal to introspection nor does introspection come with a phenomenology of its own.

My primary goal has been to refute the thesis that consciousness can be explained in terms of introspection, the thesis of introspectionism. First I discussed some reasons that might lead one to accept introspectionism. I argued that belief that introspection might shed some light on consciousness springs primarily from one's failure to distinguish one's being conscious from one's being conscious 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.

At the bottom of the level confusion between consciousness and the consciousness of consciousness there is a correct insight, viz., that consciousness has built into it a characteristic relation to an apprehending mind. But this correct insight is lost once it is subjected to the mistaken interpretation that it typically receives.

Certain traditional views about the unique kind of cognitive access that we supposedly enjoy to our qualitative mental states strongly suggest such mistaken interpretations of the relation of consciousness to an apprehending mind. This feature of consciousness is then taken to consist in the fact that the experiencing subject invariably believes or knows himself to be in a qualitative mental state just in case he is in this qualitative state.

The belief thesis—the thesis that our conscious mental states are invariably 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 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 having of which constitutes your being conscious of the corresponding qualitative state) the subject must be able to become introspectively aware of his qualitative states. Thus the belief thesis makes plausible the claim that explaining consciousness amounts to explaining introspection.

In criticizing the belief thesis I argued that having a qualitative belief is neither sufficient (the thesis of detached qualitative beliefs) nor necessary (the thesis of detached qualia) for being in a particular qualitative 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.

Up to this point the argument of this Chapter was directed against that thesis that consciousness can be explained by an appeal to introspection or that introspection can account for the production of consciousness. In the final section of this Chapter I argued for the thesis that introspection does not have a phenomenology of its own. This completed the case against linking introspection and consciousness in any way: introspection neither produces or explains qualitative mental phenomena nor does it have a phenomenology of its own. In short, introspection and consciousness have nothing to do with each other. Introspection and consciousness are not inseparably linked to each other; on the contrary, I have tried to show that the assumption that consciousness is to be explained in terms of introspection is entirely gratuitous.

CHAPTER3

HOW TO BUILD AN IMITATION MAN: OSCAR THE UNCONSCIOUS INTROSPECTOR

In the previous Chapter I tried to drive a wedge between consciousness and introspection. My main goal was to discredit the thesis of introspectionism or, more particularly, the hierarchical theory of consciousness. I pursued this goal by asking the question: What reasons are there for placing any credence in the introspectionist approach to consciousness. I argued that introspectionism arises out of a mistaken interpretation of the dimly grasped truth that consciousness has built into it a directedness upon an apprehending mind. By interpreting this directedness in terms of the belief thesis (or one of the stronger knowledge theses) one is naturally moved towards an introspectionist account of consciousness. Thus my main effort was directed at attacking the belief thesis in an attempt to rob introspectionism of its foremost motivating force.

In this Chapter I shall continue to pursue the goal of discrediting the introspectionism. But I shall proceed in a different manner. Instead of talking about introspectionism in general, I shall 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 shall try to show that Pollock's introspectionist construction does not suffice to serve up 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 in 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 then points out that qualia and the feeling of qualia become entirely unmysterious if we identify them with the outputs of certain introspective sensors that are part of our cognitive architecture. What I shall try to show in this Chapter is that we should resist this identification. I shall argue that we have good reasons for believing that a being that instantiated Pollock's hierarchical theory of consciousness might lack all qualitative consciousness. Therefore I conclude that (Pollock's version) of introspectionism fails as a theory of consciousness.

POLLOCK'S OSCAR AS IMITATION MAN

In his book *How to Build a Person* [Pollock, 1989] Pollock really does tell us how to build a person—surprisingly enough the catchy title is not the least bit misleading. Pollock is currently engaged in the project of building an artificial person—Oscar—and his book is both a report on the progress he has made and a sketch of the philosophical underpinnings of this venture. 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, p. ix-x] I shall focus on what seems to me to be the most problematical 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.³⁹ Borrowing a term from Keith Campbell I can put my worry by saying that Oscar will be an imitation man; i.e., a being very much like us except that

instead of feeling a pain when he burns his finger or breaks his toe, 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 or broken his toe, as the case may be. [Campbell, 1984, p. 100]

The charge I shall level against Oscar is not exclusively concerned with pain qualia. I shall 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 shall have to sketch a fairly detailed account of Oscar’s philosophical underpinnings. To this task I now turn.

³⁹ 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.

RATIONAL FUNCTIONALISM AND CONSCIOUSNESS

LINKING RATIONALITY AND CONSCIOUSNESS

Consciousness poses a difficult problem for functionalism. The attempts do deal with this problem span the whole range from Daniel Dennett's eliminativist advice that it is best "to declare that there simply are no qualia at all" [Dennett, 1988, p. 44]⁴⁰ to Pollock's view that functionalism *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 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

⁴⁰ See also [Rey, 1983] in [Davidson, et al., 1983, pp. 1-39].

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 who experienced nothing, Thomas 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, p. 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, p. 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, p. 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.*

THE ROLE OF CONSCIOUSNESS IN THE HUMAN RATIONAL ARCHITECTURE

Why does 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 do defeasible reasoning; to do that we must be able to keep track of the reasoning process that lead 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 “inner sense,” 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. Our inner sense “endows” [Pollock, 1989, p. 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 how possessing the rational

architecture we do leads us to enjoy the characteristic kind of consciousness that makes it like something to be us.

HOW TO BUILD CONSCIOUSNESS INTO OSCAR

FITTING OSCAR WITH INTROSPECTIVE SENSORS

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 behavioristic 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 into the Oscar machine. Some of these second-level sensors (the sensor sensors) detect activity in Oscar's first level sensors; others (the thought sensors) monitor occurrences that we may call Oscar's thoughts. We shall call the first-level sensors perceptual sensors; second-level sensors are introspective sensors. Thus Oscar turns into an introspective 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 sen-

sor-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.

FITTING OSCAR WITH SECOND LEVEL INTROSPECTIVE SENSORS

Have we switched on Oscar's inner light by giving him second-level sensors? Not yet, it seems. There is more to introspection than we have built into Oscar so far. As I understand Pollock, the phenomenal glow will not arise unless we furnish Oscar with another level of sensors: second order introspective sensors that sense the output of the introspective sensors. This interpretation is suggested by the following passage:

...qualitative feel is the output of the introspective sensors. One may protest that the "feel" of the qualitative feel is being left out of this picture...But let us be careful to 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. ...In essence, my proposal for understanding the phenomenal feel of a sensation consists of distinguishing between the sensation and the feel of the sensation. [Pollock, 1989, p. 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 white piece of paper:

- (1) The white sensor operates and produces what we might call *a perception of white*.
- (2) A first order introspective sensor operates and produces *the having of a qualitative feel, a white quale*.
- (3) A second order introspective sensor operates and produces *the experiencing of the qualitative feel, the experience of the white quale*.

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 the process. I think that Pollock holds that nothing less than level (3) will guarantee a fully conscious perceptual experience. Consider level (1) in isolation: the white perception, the output of the perceptual sensor, does not bring any qualitative aspects into play. When it occurs, unperceived 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 white sensor. A certain mental state is a white perception solely in virtue of the role it plays in a particular rational architecture; that it have or that its owner perceive it as having a certain phenomenal property is not required. Adding level (2) to the operation of level (1) seems to actually change the phenomenon: being perceived by a first level introspective sensor endows the white perception with the phenomenal property of whiteness. The white perception now has a white quale. For one who holds that qualia just are certain physical features of neurological (or electronic) states, it is obvious that such a quale can exist in a person without being noticed by the person, in which case the person will not be experiencing a phenomenal feel. To actually experience the feel/quale, it has to be perceived. That is what level (3) adds to the picture.

Directing a second-level introspective sensor upon the output of the first level introspective sensor (the white quale), makes Oscar experience the quale which he (already) has. At this level he attains consciousness; i.e. at this level it begins to be like something quite characteristic for him to look at a white piece of paper. If Oscar's three levels of sensors are operating properly it will be like something to be Oscar.

D O U B T I N G T H A T O S C A R I S C O N S C I O U S

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. In this section I shall present my reasons for doubting that Oscar is conscious. I shall 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. Moreover there exist reasons that make it reasonable to accept the possibility that all the occurrences that underwrite Oscar's reasoning do in fact take place in perfect inner darkness. Therefore I conclude that it is reasonable to assume that Oscar is unconscious.

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.⁴¹ 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, p. 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, p. 13]

As one critic notes, the most striking feature of Armstrong's characterization of perception surely is his "disregard for *sentience*." [Kneale, 1969, p. 299] 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. This same line of thought is pushed a bit further by Paul

⁴¹ Pollock notes that his account of consciousness is similar to that of D.M. Armstrong. I shall therefore take the liberty to use some of Armstrong's notions in the ensuing discussion of Pollock's theory.

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, p. 15]

I take it that Churchland uses the term “sensation” to refer to the experienced feel of a perceptual episode. I shall 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. There could be perception without sensation.

We now know that the possibility that was adumbrated in Armstrong's characterization of perception and that Churchland pointed out explicitly 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.”⁴² 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).

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: its capacity to explain this baffling phenomenon he takes to be a showcase of his theory's power.

I shall, therefore, begin by briefly describing Pollock's explanation of blindsight. I shall then present some empirical evidence (concerning the

⁴² H.D. Mellor, quoted in [Natsoulas, 1982, p. 91]. For a book-length study of blindsight see [Weiskrantz, 1986].

capacities of blindsighted persons) that is inconsistent with Pollock's explanation of blindsight. I shall argue that Pollock gets caught in a dilemma from which I do not know how to free him. Finally I shall 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.

POLLOCK'S ANALYSIS OF BLINDSIGHT

According to Pollock 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. Pollock's brief but clear discussion of the phenomenon goes as follows:

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, p. 16]⁴³

⁴³ 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.

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 Pollock'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 shall now argue that this consequence of Pollock's account is incompatible with certain observations about the capacities of blindsighted subjects.

A CONSEQUENCE OF POLLOCK'S ANALYSIS: NO LEARNING WITHOUT INTROSPECTION

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 an 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, p. 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 (that 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 shall rely on in the following discussion.

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, p. 682]⁴⁴

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, p. 37-38] 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, p. 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, p. 380-382]

Though somewhat cursory this report about some of the work that has been done with blindsighted subjects should suffice to convince us that

⁴⁴ For more on Helen see [Humphrey, 1974] and [Humphrey and Weiskrantz, 1967].

blindsight subjects are capable of being conditioned to perform better in tasks of visual discrimination. I shall now argue that Pollock's account of blindsight is not compatible with the fact that blindsighted subjects have this kind of learning power.

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 the functional organization that allows it to be conditionable. This is so because 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 theory 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 at all.

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 shall 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.

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 shall therefore propose a different way of thinking about the phenomenon of blindsight. When viewed from this angle, consideration of the phenomenon of blindsight affords one a reason for doubting that Oscar has or experiences any qualitative phenomena whatsoever.

Oscar has three levels of sensors: perceptual sensors, and first and second order introspective sensors. What we have here is three layers of perception stacked on top of each other. A primitive Oscar model that had only the first level perceptual sensors would perceive the environment without sensing it. 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 shall try to show is that adding more perceptual 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 external and internal perception without corresponding sensation, 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 perception. And if not a single one of Oscar's three perceptual layers spawns qualia, then surely Oscar as a whole is devoid of qualia too. Note that this impairment would in no way in-

terfere 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 level 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.*

AN OBJECTION TO THE ARGUMENT FROM BLINDSIGHT

It must be granted that each one of Oscar's three layers of sensors, when taken in 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.

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 level introspective sensors he discriminates between the outputs of his perceptual sensors; and with his second level sensors he discriminates between the outputs of his perceptual and his first level 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 shall 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 the more general conclusion of 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 the mental events are physical events. [Pollock, 1989, p. 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, p. 14]

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, p. 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.⁴⁵

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 shall now turn to a discussion of reasons that should incline one to take the possibility of an unconscious Oscar seriously.

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

⁴⁵ 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, p. 237].

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 shall have to argue that the envisaged scenario is a possibility that is to be taken seriously.

TWO ARGUMENTS FOR TAKING OSCAR'S BLINDSIGHT SERIOUSLY

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, p. 81]⁴⁶ should make blindsighted Oscar more attractive to us than his conscious brother. Both model human rationality, but the blindsighted one relies on less dubious means in doing so.

⁴⁶ See also [Kihlstrom, 1987].

THE ARGUMENT FROM MYSTERY

Sticking to Chisholm's advice: "try to keep mystery to a minimum," [Chisholm, 1983] 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, p. 176]⁴⁷

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, p. 82] one still can be puzzled by the question how a physical occurrence, which seems to have nothing in common with an expe-

⁴⁷ 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, pp. 795-796] But there are also those who deny the existence of the mystery outright. See, for example, [Lycan, 1987, pp. 50-52 and 76-77].

rienced feel, can manifest itself as the having of a subjective feel.⁴⁸ 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.⁴⁹

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. The 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.

⁴⁸ Cf. also [Maudlin, 1989, p. 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 or magnitude."

⁴⁹ An acute awareness of this mystery has lead 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. Most recently this mystery has lead Colin McGinn [McGinn, 1989] to pronounce the mind-body problem insoluble.

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 neither 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.

TWO ARGUMENTS AGAINST TAKING OSCAR'S BLINDSIGHT SERIOUSLY

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

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 blind-sight 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 problematical 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 licence 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.⁵⁰

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 licences 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, p. 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.

⁵⁰ 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.

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. [Cummins, 1984, p. 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.

I think that this experiment is too good to be true. Roughly, my objection goes like 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 that the subjective feel stays the same. 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*. Let us name the complex A&B *Natural*; and the complex A&C *Artificial*. Because they share 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, as I maintain, 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 is refuted.

But this objection won't stand up. 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 will become obvious that there exists no one person who experiences an unsettling color-qualia inversion upon flipping the switch. Natural would experience the meadow in which A is situated as looking just the way meadows always look; and, after the switch was flipped, Artificial would not notice anything peculiar about the color of the grass. Comparatively speaking, both of them would have been 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 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 if we pass control 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?

By parallel reasoning we will get the result that nobody will notice anything even in the case in which Oscar is unconscious. While B controls A, everything will look normal to Natural. And if the control is passed to C, then Artificial will not notice anything for he is totally unconscious.

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.

CONCLUSION

Nagel has written that “consciousness is what makes the mind-body problem really intractable.” [Nagel, 1974, p. 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, p. 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 not conscious; 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 consciousness as the cream on the cake of mentality I can summarize my conclusion by saying that Oscar has the cake without the cream.

This concludes the case against introspectionism. I have argued that the main reason for accepting introspectionism does not stand up to scrutiny (Chapter 2); and I have tried to show that the best extant introspectionist theory of consciousness does not hold what it promises (Chapter 3). 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 (this is true even for our mental lives), we shall be well advised to explore other possibilities in the meantime.

CHAPTER 4 QUALIA DE-MENTED: THE CASE FOR NONMENTAL QUALIA

GIVING UP THE ASSUMPTION THAT QUALIA ARE MENTAL

To be conscious is to have qualia. In order to unpack this slogan we have to address at least two questions: What are qualia? and What is it to have qualia? Chapters 2 and 3 addressed the second of these questions. Introspectionism is the prevailing theory of what it is to have qualia. In the two preceding chapters I stated my reasons for rejecting introspectionism. In Chapter 3 I focussed on Pollock's specific version of introspectionism and attempted to show how it fails as an account of qualitative consciousness. In Chapter 2 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. That is, in developing the anti-introspectionist argument of the last two chapters I allowed myself to rely on a largely inarticulate 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 strong assumptions about the nature of qualia. One such assumption about qualia that remained unquestioned throughout all the many turns of the anti-introspectionist argument was the assumption that qualia exist in the mind, that qualia qualify mental items.

This assumption was shared by all parties of the dispute. Introspectionism seems to presuppose the view that qualia are in the mind—how else could

introspection be a means of one's getting at them? And in my critique of introspectionism I too made the assumption of the mentality of qualia. The assumption played a particularly prominent role in my 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.

But the assumption of the mentality of qualia has not gone unchallenged. The view that qualia do not qualify mental items, i.e., that qualia are nonmental in nature has found many defenders.⁵¹ I call this thesis the *thesis of de-mented qualia*.⁵² Often, but not always, the thesis of de-mented qualia is supplemented by the claim that the nonmental qualia are properties of external physical objects. On this view qualia are located on the object with which we become acquainted in perception. The tomato's phenomenal redness, for example, is quite literally located on the nonmental surface of the tomato. In confronting the issue of the location of qualia the discussion takes an important turn. By raising the question whether qualia might be nonmental we have switched from the second to the first of the two questions listed above. Determining whether qualia are mental or

⁵¹ Not every reductive account of qualia makes qualia into nonmental properties. One can hold that qualia are identical physical properties of those physical states that are mental states. But if one takes qualia to be properties of mind independent objects, like actual (and possible) tomatoes, one treats qualia as nonmental properties. I find this view in [Armstrong and Malcolm, 1984], [Lycan, 1987], [Graham and Stephens, 1985], [Harman, 1990].

⁵² Since the term "de-mented" harbors an irresistible and crashingly obvious temptation to pun, I shall not attempt to resist it: What I shall try to convince you of (in the next chapter) is that the thesis of de-mented qualia is a demented thesis of qualia.

nonmental constitutes an important part of the answer to the question: What are qualia?

In this Chapter I introduce the thesis of de-mented qualia by way of a critique of the First Appeal to Phenomenology and I shall present and evaluate (part of) the *prima facie* impressive evidence for the thesis of the nonmental nature of qualia. I go on to ask a number of questions that the introduction of nonmental qualia forces upon us. The point of asking these questions is to get a clear sense of the repercussions that the introduction of nonmental qualia creates for introspectionism on the one hand, and for my own project of accounting for consciousness in terms of the having of qualia, on the other. The first two questions concern the bearing that the introduction of de-mented qualia has upon introspectionism. Since I have argued that the belief thesis is the main reason for introspectionism my first question will be: How does the belief thesis fare in the light of de-mented qualia? But as it is possible to be an introspectionist without basing introspectionism on the belief thesis, I then go on to ask: How does introspectionism fare in the light of de-mented qualia? I shall argue that both the belief thesis and introspectionism fare badly if qualia are nonmental. This means that the thesis of de-mented qualia gives us another reason for rejecting introspectionism.

Then I go on to ask two questions concerning the bearing of de-mented qualia on my own project of providing an account of consciousness in terms of the having of qualia. The third question I raise is this: Does the move to nonmental qualia rob the qualophile of his subject matter? I answer that the qualophile has nothing to fear. Lastly I shall ask how de-menting qualia affects the chances of one's providing an interesting account of what it is to have qualia.

Here, finally, we encounter an untoward consequence of de-menting qualia. I find myself incapable to say anything illuminating about what it is to have de-mented qualia. Thus the thesis of de-mented qualia is a mixed blessing. On the positive side it helps refute introspectionism together with its traditional underpinnings, viz., the belief thesis. On the negative side it keeps me from carrying out my own project. But all is not lost. In the next chapter I hope to show that the thesis of de-mented qualia is false. This result defeats the anti-introspectoinist argument; but, much more importantly, it opens up a possibility for an intelligible account of what it is to have qualia.

CRITIQUE OF THE FIRST APPEAL TO PHENOMENOLOGY

The first appeal to phenomenology supposedly established the existence of detached mental qualia. The thesis of de-mented qualia directly contradicts this phenomenological finding. Support for the thesis of de-mented qualia comes from two sources. First, it has been argued that the whole idea that we have qualitative mental states, a phenomenology, is mistaken; there just is no such phenomenon. This is the *thesis of the absence of phenomenology*. Second, it has been argued that the qualitative phenomena exist, but that there is nothing mental about them. These phenomenal qualities are not properly attributed to the mind; they are located “outside,” where they are to be identified with certain physical properties of (nonmental) physical objects. This is the *thesis of the relocation of phenomenology*. The absence thesis as well as the relocation thesis entail that qualia are nonmental. The absence thesis denies that there are qualia and therefore denies that qualia are the properties of mental items. By moving the

qualia out of the mind and onto the mind's nonmental object the relocation thesis guarantees that qualia are nonmental. According to both of these theses the first phenomenological verdict is false. The purported discovery of solitary or detached qualitative mental phenomena must be mistaken because there are no such phenomena in the mind. Thus we must neutralize the absence thesis and the relocation thesis if the appeal to phenomenology is to be accorded any weight. But first we need a fuller statement of the two theses.

THE ABSENCE OF PHENOMENOLOGY

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, p. 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, sound, 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, p. 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, p. 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, p. 95]

This, then, is the thesis of the absence of phenomenology: our inner lives consist of nothing but judgements. There is in the inner theater no room for solitary qualitative mental phenomena. How does this thesis bear on the question that Dennett knows how to ask so evocatively: “Is that special inner light really turned on, or is there nothing but darkness inside?” [Dennett, 1987a, p. 164] Dennett’s exposition of the thesis of the absence of phenomenology 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.

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, and etc. Discussing the experience of phosphenes, for example, 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, p. 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.⁵³ This account suggests answers to two 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.

If this is a correct interpretation of the thesis of the absence of phenomenology, then the absence thesis is virtually indistinguishable from the relocation thesis, to the discussion of which I now turn.

THE RELOCATION OF PHENOMENOLOGY

The proponents of the thesis of the absence of phenomenology may, (misleadingly, I believe) encourage the impression that they want to deny that we become aware of a rich phenomenology in perception. That perception acquaints us with a multitude of qualitative phenomena is held fast by the proponents of the relocation thesis. But they argue that there is nothing mental about these qualities. The qualities enjoyed in perception are the qualities of the external objects we perceive. The so-called qualia are not mysterious features of the mind but simply properties of the perceived object. We are dealing here with a

⁵³ 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 relocation thesis.

relocation of qualities: the qualities that were, mistakenly, taken to qualify the perceiving are now discovered to adorn the perceived. The qualities are moved out of the mind and onto its object. Once perceiving is thus understood, perception itself appears as devoid of all intrinsic phenomenal qualities. Thus the first appeal to phenomenology is undermined: the phenomenological examination of our experience cannot establish the existence of detached qualitative mental states because the qualia belong to the perceived external object; not to the mental state that is our perceiving of it. The appeal to phenomenology can, therefore, be accorded no weight in the assessment of the thesis of detached qualia.

SUPPORT FOR THE RELOCATION THESIS

Support for the relocation thesis has to come in two parts, one negative the other positive. On the one hand, we need to be given reasons that allow us to discard what seem to be crushing objections against the relocation thesis. On the other hand, the proponent of the relocation thesis needs to provide positive reasons for accepting the relocation thesis.

The two most prominent objections against the relocation thesis are the argument from illusion and the Democritean argument from atomism. As long as these objections stand, the relocation thesis falls. Both objections are directed against the claim that the perceived physical object is a promising new home for the phenomenal qualities. First, it has been argued that there are cases in which we experience phenomenal properties in the absence of any physical object that could plausibly be said to bear them. This is the case on which the argument from illusion focuses. Second, it has been argued that the relocationist thesis is

false even in those instances in which the relocationist manages to identify a physical object to bear the phenomenal properties of experience. For according to the best extant accounts of the nature of physical objects they are not the kind of thing that can have phenomenal properties. This is the claim supported by the argument from atomism. Successfully dealing with these two objections constitutes the negative support for the thesis of relocation. No matter how good the positive arguments for the thesis may be—as long as these two objections have not been silenced, the relocation thesis cannot get off the ground.

Positive support for the relocation thesis flows from two sources. First, it is alleged that the friend of qualia cannot provide a home for phenomenal qualities in the mind. Relocating these properties is, therefore, the only way to save them from elimination. Second, the champions of the relocation thesis hold that the testimony of experience shows the relocation thesis to be true.

In this chapter I shall confine myself to a discussion of the positive reasons for accepting the relocation thesis. The discussion of the two objections against the relocation thesis will be postponed until the next Chapter.

NO MENTAL HOME FOR QUALIA

The first positive reason offered in support of the relocation thesis 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 on the object of the experience, not in the experience of the 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? It meant 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 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 physical object to which the mind is perceptually related appears to be the most likely bearer of the phenomenal properties.⁵⁴

How should one reply to this challenge? The most straightforward reply would consist in a defence of sense-data. I do not want to pursue this route here, though I am not convinced of its hopelessness. Instead I shall attempt to raise problems for the claim that the physical object can be and is the home of the quale. As the next Chapter's discussion of the argument from illusion and the argument from atomism will show, establishing that the mental object cannot serve as the home for the quale falls far short of establishing that the physical

⁵⁴ Does not the relocationist present us with a false dichotomy: qualia are borne either by sense data or by the external physical objects? Where does this leave the identity theorist and the functionalist who want to hold that the qualia are properties of states of the perceiving subject's brain? This is where the second positive argument for the relocationist's position comes into play: experience teaches that the phenomenal properties belong to the perceived object rather than the neurophysiological state which is my perceiving of the object. Thus, if qualia are to be relocated their new bearers must be the external physical object (the tomato) rather than the neural/functional state that is the seeing of the tomato.

But the view that qualia are properties of neural states will not be passed over in complete silence. As the next chapter will make clear, the discussion of the relocation thesis will afford us a perspective from which it appears that moving the qualia into the brain is nearly as problematical as the attempt to move them all the way out onto the external object.

object is the home of the quale. The move from "the quale is not in the mind" to "therefore the quale must qualify the physical object" is far too quick. That is, the relocationist may have good reasons for assuming that the mind is not a good place for the quale; but, as we shall see, there are also good reasons for assuming that the object to which the mind is related is not a good place for the quale either. It appears that we get stuck with the vexing result that the nonmental world is no more hospitable to qualia than the mind is. In the light of this paradoxical conclusion we may have to back and reassess our arguments against the mentality of qualia.

THE SECOND APPEAL TO PHENOMENOLOGY

The second positive reason in support of the relocation thesis is a simple but powerful appeal to experience. To the advocates of the relocation thesis the truth of their thesis seems obvious: you just see that it is true every time you open your eyes:

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. It is not an indubitable given to be respected at all costs. There are no indubitable givens. But it is a plain deliverance of perception, and it is plausible that a true theory of perception will uphold this deliverance. [Armstrong and Malcolm, 1984, p. 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, p. 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, p. 39]

Prima facie the appeal to the experiential validation of the relocation thesis does seem highly plausible. But further consideration of the issue will, I hope, erode this assurance.

THE INCONCLUSIVENESS OF THE APPEAL TO PHENOMENOLOGY

I shall take a remark that has been attributed to Wittgenstein as my 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?" [Anscombe, 1959, see p. 151][Stoppard, 1972, p. 75]

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 a geocentric world would. 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 alternative possibilities are experientially indistinguishable. Using and extending Harman's colorful terminology, the matter can be put as follows: if you were experiencing "mental paint" it would look no different to you than if you were experiencing "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 seems to favor one of the two hypotheses over the other. But once we realize that the data we should expect to obtain on the competing hypothesis would be

indistinguishable from the data we currently possess, the inference from the data in our possession to the one hypothesis is defeated. On the other hand we have two metaphysical hypotheses and a set of experiential data that strongly seems to favor one of the two hypotheses over the other. But once we realize that the data we would have to expect on the competing hypothesis would be indistinguishable from the data we currently possess, the inference from the data in our possession to the one hypothesis is defeated. Here is how the elements of the analogy line up:

<i>The astronomical example</i>		<i>The metaphysical example</i>	
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 would be 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 would be 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 earth-bound creatures: to us as earth-bound creatures it will always look as though our planet were stationary and the sun were revolving around it. The appeal to the “obvious” explanation is no more successful in the metaphysical example than in the astronomical example. What is the correct explanation of our bias to interpret our experiential data as supporting the “physical paint” account of colors? 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 relocation thesis. Up to this last step of the analogy 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. For 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 relocation thesis is simply a “perceptual deliverance”; to undermine the claim that to deny the relocation thesis is “counter to ordinary visual experience.” 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.⁵⁵

This result may lead the relocation theorist to reevaluate his initial acceptance of the analogy. Most likely it will be the third step which will appear problematical 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

⁵⁵ This result is not entirely welcome. Since I initially based my endorsement of the mental paint story on an appeal to phenomenology, my own case is also severely weakened by this result.

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 relocation thesis have a tendency to believe that if the mental paint story were correct we should experience something like a flat painting in our heads.⁵⁶ But surely we have no reason to accept this view. Consideration of the following possibility may help to better 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 done very well, you would stop seeing two little flat screens in front of your eyes. Instead you would start to see a full-featured three-dimensional world in front of you.⁵⁷ 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

⁵⁶ 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 (1713), 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, 1954 (1932), pp. vii-viii].

⁵⁷ For a discussion of the scientific feasibility of such goggles see Hans Moravec’s discussion of “magic glasses.” [Moravec, 1988, chapter 3]

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 focussing 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 flat 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 fall short of a proof of 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 relocation thesis make much. The probative force of the second appeal to phenomenology has suffered a severe setback.

But a high price had to be paid to score this limited point against the relocationist. My own earlier appeal to phenomenology has suffered a comparable 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 the next Chapter I shall try to show that there are other reasons for believing that the relocation thesis is false.

But at this point of the debate both the mentality as well as the nonmentality of qualia are live options.

I shall take this standoff as an occasion to interrupt the discussion of the relocation thesis. In the remainder of this Chapter I want to address the question: How would things look for the introspectionist, on the one hand, and for myself, on the other, if the relocation thesis were true?

THE PRICE OF DE-MENTED QUALIA

The second positive argument for the relocation thesis has proved inconclusive. Thus the relocationist will have to rely entirely on his first argument. But before I turn to a discussion of the first positive argument for the relocation thesis I want to comment on some of the broader consequences that flow from the discussion of the theses of de-mented qualia.

Brief reflection on the relevant dialectical situation will serve to indicate these broader consequences. In Chapter 2 I made the first appeal to phenomenology. Its immediate target was the belief thesis which it falsified by pointing to the existence of detached qualia. And insofar as introspectionism is grounded in the belief thesis, the mediate target of the first appeal to phenomenology was introspectionism itself. Thus the first appeal to phenomenology may be viewed as a defeater of the belief thesis and of introspectionism. In this Chapter I introduced the thesis of the nonmental nature of qualia as a defeater of the first appeal to phenomenology. Hence the thesis of the nonmental nature of qualia can be viewed as a defeater of my defeater for the belief thesis and for introspectionism. That is, if the thesis of nonmental qualia is correct, the belief thesis and introspectionism are epistemically in the clear. It

would seem, then, that one who is committed to the belief thesis and to introspectionism should adopt the thesis of the nonmental nature of qualia as a defeater-defeater. But I shall argue that this would be a self-defeating strategy. For the thesis of the nonmental nature of qualia is incompatible with the belief thesis as well as with introspectionism. This, then, are the first two consequences that I want to draw from the discussion of the theses of de-mented qualia. First, if the qualia are nonmental then the belief thesis is false. Second, if qualia are nonmental then introspectionism is false.

Another feature of the argument against the relocation thesis cries out for commentary. While criticizing the second appeal to phenomenology I admitted that an appeal to phenomenology cannot be an arbiter between the mental paint and the physical paint story. But recall that my endorsement of the mental paint story rested, primarily, on the first appeal to phenomenology. It seems, then, that my endorsement of the mental paint story is unfounded. A critic might argue that this spells doom for my project. For if the phenomenal properties are kicked out of the mind, the qualophile is deprived of his subject matter. This consideration gives rise to the third consequence that I want to draw from the discussion of the theses of de-mented qualia: contrary to appearances the qualophile has nothing to fear from the relocation move.

I end on a less positive note. Reflecting on the question what it would take to have de-mented relocated qualia, I find myself unable to provide a satisfactory answer. This falls far short of showing that such an account cannot be produced. But as far as I am concerned, it puts an end to my goal of explaining consciousness in terms of the having of qualia.

DE-MENTED QUALIA AND THE BELIEF THESIS

What would follow if the relocation thesis were correct after all and phenomenal properties were properties of nonmental objects? Would this make the belief thesis any more plausible? I shall now argue that this is not so.

The relocation thesis may seem to pose a threat to my argument because it appears to undermine my case against the belief thesis. If the relocation thesis 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 relocation thesis would seem to strengthen the belief thesis by defeating my defeater for the belief thesis. This consideration for believing that the relocation thesis supports the belief thesis is correct as far as it goes. But it stops short of fully appreciating the epistemic punch of the relocation thesis. For further reflection will show that the truth of the relocation thesis is itself a defeater of the belief thesis.⁵⁸ It follows that I need not lose my composure if the relocation thesis should be true.

The case would be absolutely clear if we were dealing with a stronger interpretation of the belief thesis. Were we to interpret the belief thesis as, for example, the simple identity thesis—the thesis that holds that being in a particular conscious state and believing that you are in this state are literally the same

⁵⁸ This last claim may seem terribly obvious. For the nonmental qualia that grace the surfaces of external objects must surely count as detached qualia? And detached qualia falsify the belief thesis. As I try to show in the text, this assessment is correct as long as we are dealing with strong versions of the belief thesis. I hope to show that the relation of the relocation thesis and the belief thesis become much more interesting and much harder to assess as we encounter ever weaker versions of the belief thesis.

state—the case would be straightforward. If qualia are surface properties of physical objects then, surely, the having of a quale cannot be identical with a qualitative belief about it. To deny this would commit one to the view that qualitative beliefs are surface properties of things like tomatoes, which is absurd. Things turn out no better if we start with qualitative beliefs. Presumably a qualitative belief is identical with the qualitative state it is about we get the result that the qualitative state in question is a mental state. In other words, the quale in question ends up being “unrelocated;” that is, it qualifies a mental state. Thus the relocationist cannot consistently help himself to a strong version of the belief thesis.⁵⁹

Matters are less clear-cut in the case of the weak version of the belief thesis that we are looking at—the thesis that holds that one’s having the relevant qualitative belief is a necessary condition of one’s being in the corresponding qualitative state. As stated the belief thesis requires that the subject having the qualitative state and the subject having the qualitative belief be the *same* subject. This has the untoward consequence that only things that are capable of belief can have (states with) phenomenal properties. A ripe tomato, for example, will be red only if it believes itself to be red. Since it can’t do the latter, it won’t be the former, which is false. Therefore the belief thesis is false if the relocation thesis is true.

One might try to remedy this problem by arguing that the phenomenal redness is not a feature of the tomato, but, instead, a feature of the neural state of

⁵⁹ But these considerations are of limited importance because we decided early on in chapter 2 to discuss a weak version of the belief thesis. One might, however, construe these considerations as one more reason for choosing a weak version of the belief thesis as the target for our critique.

the one who sees the tomato. Somewhat facetiously one might name this version of the belief thesis the *colored brain thesis*. Since it may be assumed that one who is capable of seeing a tomato is also capable of believing that he sees a tomato the requirement that it be the same subject who has the qualitative state and the corresponding qualitative belief no longer poses a problem. But this move flies in the face of the relocation thesis. The idea was to get the phenomenal properties out of the mind and onto the mind's object. By moving the phenomenal properties out of the mind and onto the (physical) brain states we have, at best, reached a halfway house. The relocationist wants to move them all the way out, onto the object seen (which usually will *not* be part of your brain). Disregarding the question whether the present proposal captures the spirit of the relocation thesis, it seems that that it is plainly wrong. For if our brain states are the bearers of the phenomenal properties of which we are aware, then our brain should be red, green, loud, sour, etc., which it is not.⁶⁰

A more promising attempt to repair the belief thesis results if we relax the requirement that the subject having the qualitative state and the subject having the qualitative belief be the same subject. This maneuver allows us to peel the qualitative beliefs about phenomenal redness off the tomato's surface and to locate them in a subject (other than the tomato) that is capable of having beliefs. This proposal—the *relaxed belief thesis*—seems to be true to the spirit of the

⁶⁰ 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 colored brain thesis I should say that it is not at all obvious to me that the more sophisticated versions of this thesis have been refuted. Bertrand Russell, for example, defended a theory that is closely similar to the one discussed above. He argued that all we ever see are (constituents of parts of) our own brains. And to the standard objection that the neurosurgeon does not see my brain turning green when I look at grass he replied that this was so because the neurosurgeon was only observing his own brain, not mine.

relocation thesis: the phenomenal redness is out there on the surface of the physical tomato, where it belongs. And, at the same time, it seems that we preserved something very close to the original version of the belief thesis: the tomato's redness depends upon its giving rise, in an appropriately situated subject, to the qualitative belief that the subject is being appeared redly to. Is this, then, a way to reconcile the truth of the relocation thesis with the truth of (one version of) the belief thesis? If so, I must retract my claim that the belief thesis is false if the relocation thesis is true.

But the relocationist cannot rest content with the relaxed belief thesis. The problem with this version of the belief thesis does not pertain to *where* it locates the phenomenal properties but to *what* it locates there. According to the relaxed belief thesis the secondary properties are thoroughly subjective, flickering in and out of existence depending on whether they do or don't happen to give rise to a corresponding qualitative belief in an appropriate subject. Surely such fickle and mind-dependent qualities are unacceptable for the relocation theorist. For they have none of the ontological respectability of the physical objects they grace.

Can we further relax the relaxed belief thesis? It would seem so. The relaxed belief thesis is easily modified to meet the problem of flickering qualia. Instead of requiring that there be a subject which actually holds the relevant qualitative belief we require merely that if there were an appropriately situated subject then it would form the relevant qualitative belief. Call this the *loose belief thesis*.

The loose belief thesis has a familiar and reassuring ring to it. It is reminiscent of a Lockean dispositionalist account of secondary properties. But if we want to stay true to the relocationist thesis we must resist the temptation to

appropriate the respectability of the Lockean account for the loose belief thesis. For the Lockean ends up locating the phenomenal redness of the tomato in the sensations (i.e. in the mind) of the tomato's (potential) beholder. The tomato itself may also, and with propriety, be called red. But by calling the tomato red, I merely ascribe a power to it: the power to produce phenomenally red sensations in an appropriately situated observer. Obviously this is a position that is unacceptable for the relocation theorist, whose goal it is to banish the phenomenal properties from the mind. [Armstrong and Malcolm, 1984, pp. 177-187]

I think that the apparent similarity between the loose belief thesis and Lockean dispositionalism is, at least in part, spurious. Thus the worry that the loose belief thesis might push the phenomenal properties back into the mind *in this particular manner* is wrong headed. But I shall now argue that the corrected understanding of the loose belief thesis holds no more promise than its Lockean misinterpretation, insofar as the reconciliation of the relocation thesis with the belief thesis is concerned.

The Lockean reading of the loose belief thesis goes wrong in suggesting that the loose belief thesis is committed to the existence of mental paint. But there is no need to interpret the loose belief thesis along these lines. The Lockean has phenomenal redness reside in sensations (in the mind). If the loose belief thesis is to be compatible with the relocation thesis the phenomenal redness must be pushed back out onto the object. This is a change which the loose belief thesis can accommodate. This version of the belief thesis does require that there be a subject who can have the qualitative belief that it is appeared redly to. But if we assume (at least for now) that one can have a qualitative belief about

phenomenal redness without having phenomenal redness instantiated in one's mind then this requirement can be satisfied without the need to postulate phenomenally red mental objects. Thus the loose belief thesis is interpretable in a manner that sidesteps the first problem raised by the Lockean reading of this thesis.

But the Lockean reading of the loose belief thesis also highlights a second feature of the loose belief thesis, viz., that the latter thesis provides a dispositional account of secondary qualities. In this respect the Lockean reading is not misleading. The loose belief thesis does furnish a dispositional account of secondary properties. The relocation thesis, on the other hand, presents secondary properties as categorical properties of physical objects. The claim that the clash between the loose belief thesis and the relocation thesis is a clash between a dispositional and a categorical conception of secondary properties can be developed along the following lines. What is it for an object to be red, according to the loose belief thesis? It is for an object to be such that it could cause a corresponding phenomenal belief in an appropriately situated observer. That is, for an object to be red is for it to be such that if there were an appropriately endowed subject present then this subject would form a certain phenomenal belief. There is, then, a conceptual connection between the statement that the tomato is red and the subjunctive conditional stated in the previous sentence. This is the basis for claiming that the relaxed version of the belief thesis provides a dispositional account of secondary qualities. On the other side we have the relocation theorist. He cannot rest content with a dispositional analysis for secondary qualities. Since he identifies secondary properties with primary properties and since primary properties are the paradigms of

categorical properties, the relocation theorist has to insist on the categorical nature of secondary properties. On this construal of secondary properties it is not constitutive of phenomenal redness that it be able to cause observers to form such and such beliefs. This clash between dispositional and categorical construals of phenomenal properties makes the loose belief thesis incompatible with the relocation thesis. Thus the truth of the relocation thesis shows the falsity of even the most watered down version of the belief thesis that we have considered.⁶¹

Neither the original belief thesis, nor the colored brain thesis, nor the relaxed belief thesis, nor even the loose version of the belief thesis are compatible with the relocation thesis. I conclude that the belief thesis and the relocation thesis are incompatible. Based on an argument from exhaustion, the case for incompatibility is not as robust as we might wish it to be. But for now the ball is in the court of the belief theorist.

In Chapter 2 I argued that the belief thesis must be given up if the mental paint scenario is assumed. Now I argued that the belief thesis must be given up if the relocation thesis is correct and the physical paint scenario is assumed. Things do not look good for the belief thesis: no matter how the dispute about the relocation thesis is resolved, the belief thesis fails one way or another. Inasmuch as I am concerned to refute the belief thesis I need not worry whether the paint is

⁶¹ One might, moreover, argue with some plausibility that the loose belief thesis is much too watered down to preserve what is important about the original version of the belief thesis. The belief thesis was motivated by its capacity to guarantee that every conscious phenomenon is accompanied by a corresponding cognition that requires the conscious subject to introspect the quale in question. This is how the belief thesis manages to marry consciousness to introspection. The loose belief thesis does not come with a similarly impressive guarantee. For it allows for unintrospected qualia. All that it can guarantee is that under certain circumstances the quale could be introspected.

mental or physical. Thus I can face the possibility of a relocationist victory with equanimity. Nothing hangs on it as far as the belief thesis is concerned.

DE-MENTED QUALIA AND INTROSPECTIONISM

The belief thesis itself is of limited interest. What interest it has it derives from its connection with introspectionism. Lets therefore turn to the question whether the relocation thesis is compatible with introspectionism.

In the previous section I argued that the relocation thesis is incompatible with the belief thesis. What bearing does this result have on the relation between the relocation thesis and introspectionism? The answer to this question will depend on how we view the connection between the belief thesis and introspectionism. The tighter the connection, the more relevant the previously reached results will be.

If one construes the process of introspecting one's qualitative states as consisting in one's acquiring the relevant qualitative beliefs then the carry-over from the previous discussion will be significant. The previous discussion has shown that if the relocation thesis is true then the belief thesis is false. That is, if qualia are nonmental then it is false that every qualitative state is invariably accompanied by a corresponding qualitative belief. Hence there are qualitative phenomena unaccompanied by the corresponding qualitative beliefs. On the present understanding of "introspection" this means that these qualitative states occur unintrospected. One's consciousness of these qualitative states will, therefore, not be explainable by appeal to their being introspected. On this construal of the notion of introspection, introspectionism cannot be offered as a complete account of qualitative consciousness.

The considerations of the last paragraph are not directly relevant to the discussion of introspectionism as it is contained in Chapter 2. For there we worked with a somewhat weaker relationship between the belief thesis and introspectionism. The belief thesis was presented as the most likely reason for one's accepting introspectionism. This kind of connection between the belief thesis and introspectionism is still strong enough to allow us to exploit the previous discussion of the belief thesis as a means to throw light on the relationship between the relocation thesis and introspectionism. The previous discussion has shown that if the relocation thesis is true then the belief thesis is false. Thus, if the relocation thesis is true one is deprived of one's main reason for accepting introspectionism. While this result falls short of proving the relocation thesis incompatible with introspectionism it does show that the truth of the relocation thesis bears negatively on the epistemic standing of introspectionism.

So much for the attempt to milk the incompatibility between the relocation thesis with the belief thesis for arguments that show that there is a tension between the relocation thesis and introspectionism. What can we say about the relationship between the relocation thesis and introspectionism without taking the detour through the belief thesis? This is not a merely academic question. For as the example of Pollock's introspectionism has shown, it is perfectly possible to be an introspectionist without relying on the support of the belief thesis.

After having gone through the tedious and ultimately unsuccessful process of progressively weakening the belief thesis in an attempt to make it consistent with the relocation thesis, it comes as a relief to discover that introspectionism does not admit of a parallel process of successive weakenings undertaken in the

interests of consistency restoration. A rather pedestrian consideration shows that all versions of introspectionism are inconsistent with the relocation thesis. If being conscious, that is, if the having of qualia is accomplished through a process of introspection, then the qualia have to be accessible to the inner sense; they have to be in the inner sense's field of vision, so to speak. To be introspectible an item has to be located where the mechanism of introspection can reach it. No matter which theory of introspection you favor, as long as it is a genuine form of *introspection* this constraint of the location of introspectible items precludes them from being situated "outside," as the relocationist would have it. This shows that if the relocation thesis is true and qualia are external to the mind, then you cannot get an introspectionist account of consciousness.

I can conclude this discussion of the compatibility of introspectionism and the relocation thesis in precisely the same way as I concluded the preceding section on the compatibility of the belief thesis and the relocation thesis. In chapters 2 and 3 I argued that introspectionism must be given up if the mental paint scenario is assumed. Now I argued that introspectionism must be given up if the relocation thesis is correct and the physical paint scenario is assumed. Things do not look good for introspectionism: no matter how the dispute about the relocation thesis is resolved, introspectionism fails one way or another. Inasmuch as I am concerned to refute introspectionism I need not worry whether the paint is mental or physical. Thus I can face the possibility of a relocationist victory with equanimity. Nothing hangs on it as far as introspectionism is concerned.

DOES THE RELOCATION THESIS ROB THE QUALOPHILE OF HIS SUBJECT MATTER?

After having struggled so hard to keep the qualia mental, I shall now ponder the possibility that I shall have to beat a retreat on this front. This is not merely frivolous flirtation with an outlandish possibility. The following consideration recommends this procedure. In the course of the discussion of the relocation thesis I was forced to admit that the appeal to phenomenology cannot decide the issue of the mentality or nonmentality of qualia. For I had to acknowledge that the mental paint scenario and the physical paint scenario are phenomenologically indistinguishable. Thus my main reason for endorsing the mental paint interpretation of phenomenal qualities is severely weakened. What follows from this reversal? I shall 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. And the critic might want to hold this view even if he is willing to grant the point that the truth of the relocation thesis does not weaken the case against the belief thesis or against introspectionism. 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 and the only slightly more important battle against introspectionism

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, p. 73]

Graham and Stephens turn what seemed like a troublesome consequence of the relocation thesis into a ingenious way of dealing with qualia-related problems. 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 then 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, then 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 relocation thesis *per se* as either posing a threat or as holding a promise for my enterprise. If the relocation thesis is false and qualia are mental then I shall do philosophy of mind. If the relocation thesis is true and qualia are nonmental then I shall have to do more than merely philosophy of mind. Shifting the qualia around does not make them go away. And as long as the qualia exist somewhere I shall 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.

WHAT IS IT TO HAVE DE-MENTED QUALIA?

It is all very fine to affect a liberal air about the nature and location of qualia. But to allow for the existence of qualia, be they inside or outside, is not all it takes to account for qualitative consciousness. After all, by moving the qualitative redness onto the surface of the tomato we did not want to grant it a qualitative consciousness of its own. The tomato's inner light did not start to shine any brighter by having a bunch of qualia dumped on its surface. From the tomato's point of view the relocationist's maneuver changes nothing: to be a tomato with qualia is just like it is to be a tomato without qualia, viz., like nothing. For there to be consciousness the qualia need to be had in the right kind of way. The tomato does not have them in the right kind of way. But if we are

to be conscious we must have the qualia in the relevant way. What, then, would it be to *have* the relocationist's de-mented qualia? How does one make these publicly available qualia one's own? How do we manage to be conscious in the relocationist's world where there is only physical paint?

The term "naive realism" has received more (and more variegated) characterizations than seems fitting for a respectable philosophical term. So I feel rather confident in saying that according to at least one version of naive realism our seeing an (external) object is simply a matter of our gaining a direct visual awareness of the object in question.⁶² The most striking feature of such an account of visual perception is that it posits no representations of any kind in the perceiver. On this picture 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 aware of the tomato's phenomenal redness.

There is a battery of standard objections to naive realist theories of perception, concerned, for the most part, with problems about illusions and

⁶² This kind of story may be, but need not, be backed up by some kind of causal account of the transaction. But since this kind of account is supposed to be applicable to us as well as to disembodied Cartesian souls the purpose of such a causal story would not be to analyze the notion of direct awareness; its purpose would be to give an account of the mechanics that happen to underlie our becoming directly aware of an object.

matters of perceptual relativity. These are genuine difficulties, and I do not want to belittle them by passing over them at this point. My biggest complaint about the naive realist account is that I find it altogether unilluminating. “Direct awareness” functions as a primitive term in the naive realist account. And it is one of those primitives that do all the work that 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.

There is a second rather ill defined worry I have about naively realistic account of consciousness. The question I want to ask is this: Where is my qualitative consciousness located on the naive realist account? My head seems to be out of the question. For the qualia are outside of it and nothing in my head or mind represents these qualia locally. Edwin B. Holt, one of the American New Realists, bravely faced up to this issue and identified a subject’s qualitative consciousness with the qualia to which the person reacts. [Holt, 1914. see especially Chapter 9] This means that your visual consciousness is, quite literally, spread out over or identical with the relocated qualia on the surfaces of the objects you see. Distinguish Holt’s claim from the true and uninteresting view that the manifold of physical objects with which vision acquaints us is extended in space. On Holt’s view the consciousness itself gets bloated to the point where it becomes spatially coextensive with its objects. I believe that Holt’s view on the location of consciousness is the correct consequence to draw from naive realism. I take that to be one more reason to be suspicious of the naively realistic account of what it is to have a nonmental quale.

These reflections do suggest that a satisfactory account of consciousness will be hard to come by within the physical paint scenario of the relocationist. Moving the qualia out of the mind and onto the mind's object makes them more respectable denizens of our ontology. But as long as the relocation move is not supplemented by a corresponding story about how our minds can reach out to the relocated qualia so as to make them our own, we shall not know what it is to have qualia, what it is to be conscious. I have not shown that such a story cannot be told. But I do not know how to tell it. Thus the truth of the relocation thesis would effectively block me from pursuing my goal of developing an account of qualitative consciousness. My critic comes out victorious, albeit for reasons he never suspected. The relocation of qualia does not rob me of my subject matter. But the relocation of qualia makes the task I set myself harder than I can handle. My equanimity concerning the relocation thesis was misplaced. With this undesirable consequence of the relocation thesis we have finally uncovered a reason for hoping that the relocation thesis is false. In the next Chapter I shall attempt to show that the relocation thesis is false. Thus we need not fear that this particular problem will keep us from formulating a satisfactory account of consciousness.

S U M M A R Y

In this Chapter I did two things. First I introduced the topic of the nonmental nature of qualia by way of considering possible objections to the first appeal to phenomenology. The thesis of the absence of phenomenology and the thesis of the relocation of phenomenology rebut the first appeal to phenomenology by holding that there are no mental qualia. The absence thesis holds that there are

no mental qualia because there are no qualia. The relocation thesis holds that there are no mental qualia because qualia are nonmental. I argued that the absence thesis collapses into the relocation thesis and proceeded to present two and to discuss one of the positive arguments for the relocation thesis. The first argument holds that qualia must be nonmental because the mind cannot accommodate qualia. The critique of this argument has been saved for the following Chapter. The second argument for the relocation thesis consisted in an appeal to phenomenology. I discussed this second appeal to phenomenology at some length and concluded that it cannot serve to establish the relocation thesis. These anti-phenomenological considerations proved to be a double edged sword. For I was forced to admit that the considerations marshalled against the second appeal to phenomenology count equally against the first appeal to phenomenology, thereby severely weakening my case for the mentality of qualia.

The remainder of the Chapter was dedicated to assessing the repercussions that the thesis of de-mented qualia has for the viability of introspectionism and my own project respectively. I argued that if the relocation thesis is true then the belief thesis is false. This supports the argument of Chapter 2. I argued that if the relocation thesis is true then introspectionism is false. This supports the argument of chapters 2 and 3. I argued that the truth of the relocation thesis does not entail that there are no qualia. This fits in with the larger project of accounting for consciousness in terms of the having of qualia. Finally I had to admit that if the relocation thesis is true I do not know how to provide an account of the having of qualia, i.e., of consciousness. This shows that I cannot proceed with my project without first refuting the relocation thesis. To this task I now turn.

CHAPTER 5

SEARCHING FOR A NONMENTAL HOME FOR QUALIA: THE CASE AGAINST NONMENTAL QUALIA

The phenomenological battle over the location of qualia has ended with a draw. The assessment of the relocation thesis will therefore have to turn on the success or failure of the relocationist's first argument for extruding qualia from the mind. Roughly put, the relocationist maintained that the qualia must be outside since there are no appropriate qualia-bearers in the mind. I shall challenge this argument by showing that the nonmental world does not tolerate qualia. That is, qualia cannot be outside since there are no appropriate qualia-bearers in the nonmental world. Thus we must conclude that the relocationist's thesis is false. I shall support this conclusion by two arguments. First, I shall show that nonveridical experience poses a formidable problem for the relocation move. In the second argument I will take on the relocation theorist in the domain of veridical experience. I shall try to show that the qualia of veridical experience cause the relocationist precisely the same kind of problems that he encountered in the case of nonveridical experience. I conclude that the relocation move is not a viable option because the nonmental world does not contain any qualia-bearers.

ILLUSIONS AND NONEXISTING INTENTIONAL OBJECTS

Illusions, hallucinations, dreams, and imaginings pose a problem for the relocation thesis. 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.

We need not now determine whether the many arguments that have been built on these various kinds of non-veridical experience (and which can, with a greater or lesser degree of justification, be lumped together under the name “the argument from illusion”) are powerful enough to establish what their proponents have, for the most part, taken them to establish: that sense data must exist and that therefore materialism is wrong. 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 imaginings—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 comes to this: in the cases of non-veridical experience he cannot provide a home for qualia.

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 (nonmental) physical object, albeit one that happens not to exist.

The defence of this position proceeds by pointing out the intentionality of sensation. Sensations are always sensations 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. No matter if 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. But as long as there is an object of any kind there is a place for the extradited phenomenal qualities. Thus non-veridical experience does not embarrass the relocationist 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, p. 841] Concerning a situation in which Alice is hallucinating a nice 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, p. 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 and Malcolm, 1984, p. 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.” In short, “The illusory qualities do not qualify anything at all.” [Armstrong and Malcolm, 1984, p. 174] A fortiori they don’t qualify anything mental like, e.g., sense data.⁶³

It appears, then, that the relocationist’s appeal to the intentionality of sensation has secured non-mental objects that can serve as bearers of phenomenal properties in cases of non-veridical experience. Thus it would seem that the first objection to the relocation thesis is overcome.

PROBLEMS WITH INTENTIONAL OBJECTS

I shall now argue that intentional objects are a poor home for qualia; especially so if they are introduced in an attempt to save the relocation thesis. Not just any notion of an intentional object will be able to do the work the relocationist needs

⁶³ William Lycan has employed the same reasoning to provide the illusory phenomenal properties with yet another pickwickian home. Not nonexistent physical objects but nonexistent sense data are the bearers of qualia. Of Leopold, who is having a green afterimage, he says: “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 after-images and other sense-data are intentional objects that do not exist? If they do not exist, then—*voilà!*—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, p. 88]

it to do. It will, therefore, be useful to remind ourselves of the kinds of constraints that an intentional object will have to satisfy in order to be of any use to the relocation theorist. First, the intentional object will have to be nonmental. For the whole point of relocating the phenomenal properties was to find nonmental bearers for them. Second, the intentional objects will have to be of a kind that can plausibly be said to have phenomenal properties. Unextended point-particles do not, for example, seem like the right kind of bearer for phenomenal qualities like redness. And third, the intentional objects will have to be such that their having phenomenal properties can explain our having the conscious experience we do. If, for example, we were to push phenomenal red into some world to which the experiencing subject bore no relation, then something's being red in that faraway world could not explain how this quality affected the subject's experience.

The first constraint on suitable intentional objects is obvious; but the last two merit further comment. The relocationist's aim in presenting us with a theory of intentional objects is to overcome the problems posed by nonveridical experience. His theory will be successful only if the intentional object, as he conceives them, are capable of providing a home for phenomenal qualities in cases in which there is no external physical object to bear them. This puts a constraint on what are possible candidates for intentional objects: they must be objects of a kind that can have or exemplify phenomenal properties. Thus a question we shall 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 shall call this the *exemplification problem*. Every relocationist theory of intentional objects will have to deal with the exemplification problem.

But solving the exemplification problem is insufficient to guarantee the success of a relocationist's intentional object account of nonveridical experience. For, presumably, the relocationist will have certain views about the relation obtaining between the qualified intentional object and the experiencing person. The experiencing person experiences the particular phenomenal qualities she does in virtue of the intentional object's having these qualities. And if the relocation theorist is naturalistically inclined, he will want to spell out the "in virtue of" locution in causal terms. Thus we arrive at a second constraint on what are 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 shall 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 interactions with experiencers? I shall call this the *causation problem*. At least all those relocationist theories that were conceived in the naturalistic spirit will have to deal with the causation problem. The exemplification problem is logically prior to the causation problem. If you fail to solve the first then you failed completely; the causation problem does not even pose itself.

A review of various proposed construals of intentional objects will make it clear that these three constraints on intentional objects are hard to satisfy. Such an argument by elimination falls short of a proof that intentional objects of the kind required by the relocationist cannot be had. But it does show that the attempt to save the relocation thesis by an appeal to intentional objects is an extremely problematical undertaking.

INTENTIONAL OBJECTS A LA (EARLY) BRENTANO

The point of the relocation thesis 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 relocationist 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 in 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, pp. 88-89]

In this brief passage 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 champion 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, p. 136]

The first two points pose no threat to the relocation thesis. For if the relocation theorist 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 problematical 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, p. 136]

Brentano provides the relocation theorist with an intentional object, but it is an object of the wrong kind. For it is a mental object and as such the relocationist cannot accept it as a suitable support for phenomenal qualities.

INTENTIONAL OBJECTS A LA MEINONG

The better part of the ensuing discussion of 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 this laudable drive towards ontological parsimony is one that the relocation theorist cannot embrace wholeheartedly. For he must make sure that he does not end up empty handed: that is, he runs the risk that after the successful elimination or reduction of intentional objects he is left with nothing that could plausibly be styled as a support for qualia. Suppose, by way of an example, that you are able to argue successfully that our intentional attitudes are really directed towards linguistic entities. What is the relocation theorist to do under these circumstances? He cannot plausibly maintain that the phenomenal qualities of experience are intrinsic properties of linguistic entities. This form of linguistic reductionism of intentional objects is therefore one that the relocation theorist must resist. Only those accounts of intentional objects will merit the relocationist's attention that construe intentional objects as nonmental objects to which phenomenal properties can plausibly be attributed. I shall now argue that the attempt to construe intentional objects as Meinongian objects—objects beyond being and nonbeing—succumbs to just this problem. The objects with which this account furnishes you cannot do the kind of work that the relocationist requires an intentional object to do.

One way to guarantee that you are not dealing with a mental object, i.e., an object that exists in the mind, 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. This insight provides the basis for a different conception of intentional objects, one that is closer to the one that Harman, Armstrong, and Lycan seem to have in mind. Nonexistence is the one feature of intentional objects upon which Harman, Armstrong, and Lycan agree. Harman and Lycan both claim that the person

suffering a non-veridical experience is related to a an object that does not exist; Armstrong calls it a non-real object. This strongly suggests a Meinongian view of intentional objects. According to Meinong the objects of our sensations (*Empfindungsgegenstände*) are “homeless objects” [Meinong, 1907, pp. 8-14]; homeless “because there would seem to be no place for [them], either in Plato’s heaven or on earth.” [Chisholm, 1973, p. 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: “[H]ow can an unreal tomato be really red?” [Price, 1954 (1932), p. 105], p. 105) Both the exemplification problem and the causation problem cry out for an answer on the Meinongian construal of intentional objects. We are owed answers to two questions: First, What is it for a homeless object to be red? Second, What good would a red homeless object 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 that 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.”⁶⁴ But I think that 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. For 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 unproblematical affair for a homeless object to be red. Can the relocation theorist rest content with the claim that nonveridical experience

⁶⁴ I think that the second part of this answer is problematical. See the next section for a discussion of the question what it is for something to be red in a non-actualized possible world.

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"?

INTENTIONAL OBJECTS AND POSSIBLE WORLDS

I have argued that the framework of possible worlds does not offer adequate interpretative resources for Meinongian intentional objects. But 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 that we arrived at independently, we shall try to fashion our notion of an intentional object along the lines that the framework of possible worlds suggests. This approach urges us 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 possible in short. Since we are now working without a preconceived view about the existential status of intentional objects, we shall stipulate that intentional objects exist in precisely the way in which objects in non-actualized possible worlds exist, whatever that may be. 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 chose. There are, of course, numerous extant accounts of possible worlds. Following David Lewis, I shall 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 (i.e., of intentional objects) will differ greatly. I shall argue that, in the end, it does not matter which branch you chose. Neither one of the resulting notions of intentional objects can do what the relocationist needs them to do. But they fail for different reasons. I shall argue that theories of intentional objects conceived in the spirit of Lewisian realism can solve the exemplification problem but fail to solve the causation problem. The ersatzist theories of intentional objects either fail to solve the exemplification problem and therefore don't even get a shot at solving the causation problem; or, if they do provide an answer to the exemplification problem, they cannot deal with the causation problem.

LEWISIAN REALISM

For the Lewisian realist all possible worlds exist in the most robust sense of the word. Possible worlds are and contain good solid physical objects, just like the ones we encounter every day. Mere possibilia are regular physical objects and

can exemplify phenomenal properties with the same ease the actual objects of this world exemplify them.⁶⁵ Since mere possibilia are, on this account, robust 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 his qualia-laden intentional objects to color his experience. So all the many worlds full of wonderfully colored and sweet smelling possibilia are in vain. Becoming a Lewisian realist about intentional objects does not advance the cause of the relocationist.

ERSATZISM

The case of the ersatzers is less straightforward, if only because their number and their disagreements are great. I shall 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 always 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

⁶⁵ The issue whether any physical objects, actual or merely possible, can exemplify phenomenal properties is an issue that will be discussed under the heading "The Argument from Atomism."

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 affairs 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). [Plantinga, 1976, p. 258] What does this setup have to offer the relocationist 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 that states of affairs of *the rat’s being pink*. But the state of affairs of *the rat’s being pink* obtains 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. For what would it be for an abstract object like a state of affairs to be phenomenally pink? That 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; neither 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 possible you just got stuck with, you might attempt to "thin out" the problematical 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 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 comfortably exemplify phenomenal pinkness. Thus it would seem that the ersatzist cannot solve the exemplification problem.

The ersatzist will impatiently reply that I have overlooked the most straightforward and obvious solution of the exemplification 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 problematical than any physical object's being pink. Thus the exemplification 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, might have been pink? The rat might, after all, also have been green (or blue, or yellow);

why, then, do I experience a pink quale rather than a green one? It would seem that the possibility that a particular grey thing might have been pink is too unspecific a state of affairs to bear the explanatory burden that the relocationist wants to place on it.

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 nonexistent objects. They are physical objects much like the ones we interact with every day. The robustness of the objects that are involved allows these theories to deal successfully with the exemplification problem. But the two positions pull apart when we look at the reasons for their incapacity 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 on this scheme of things 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.

I conclude that neither Lewisian realism nor ersatzism serve up workable notions of qualia-bearing intentional objects. Hence the relocationist is still faced with the problem of providing an intelligible notion of intentional objects that

will get him out of the problems posed by nonveridical experience. A simple appeal to possible worlds will not do.

THE ARGUMENT FROM ATOMISM AND ILLUSORY QUALIA

I have argued that nonveridical experience of the sorts that the argument from illusion focuses on poses a rather formidable problem for the relocation thesis. But perhaps nonveridical experience does not deserve so 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, therefore, 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 just that, viz., that the relocation thesis is untenable even in the cases of veridical experience and that it must, therefore, be given up.

Simply put, the problem is this: qualia don't agree with matter. The more we learn about matter, the less hospitable does it appear to phenomenal properties. 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 started with Democritus who held that "by convention color exists, by convention bitter, by convention sweet but in reality atoms and the void" [Hardin, 1988, the quotation appears on p. 59] and this 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. [Eddington, 1929, p. see p. 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, p. x] ⁶⁶

Although Eddington does not explicitly assert that the two pictures are incompatible, this view 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, p. xii] Thus the ordinary table of everyday life—"strange compound of external nature, mental imagery and inherited prejudice" [Eddington, 1929, p. xii]—loses its status as a real thing. And towards the end of his investigation Eddington informs us that "[o]ur conception of the familiar table was an illusion." [Eddington, 1929, p. 323]

Clearly Eddington's assertions could be reconstructed 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

⁶⁶ Six years before Eddington's scientific attack on the solidity of ordinary tables Bertrand Russell mounted a similar offensive against tables (and, being of a more radical temperament than Eddington, also against chairs): "To the eye or to the touch, ordinary matter appears to be continuous; our dinner-table, or the chairs on which we sit, seem to present an unbroken surface. We think that if there were too many holes the chairs would not be safe to sit on. Science, however, compels us to accept a quite different conception of what we are pleased to call "solid" matter; it is, in fact, something much more like the Irishman's definition of a net, "a number of holes tied together with pieces of string." Only it would be necessary to imagine the strings cut away until only the knots were left." [Russell, 1923, p. 1]

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 the color qualia I experience upon viewing a table on the table itself must fail. This is a sketch of the argument from atomism and how it is supposed to undermine the relocationist move.

I shall not discuss the question whether the natural philosophers from Democritus to Eddington have been right in maintaining that a scientific view of matter is incompatible with the commonsensical conception of physical objects. Due to Wilfrid Sellars's 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 are now in existence.⁶⁷ 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. Instead I shall focus on one aspect of a philosophical thesis that attempts to defend the relocationist thesis in what I take to be a most original way. By that I mean that this defence relies on a strategy that was undreamed of by the Greek atomists, the modern corpuscularians and twentieth century physicists such as Eddington. One thing that stood fast for all 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

⁶⁷ See, for example, [Sellars, 1962]; [Aune, 1967, pp. 172-175]; [Hilbert, 1987, p. 43-46]; [Lycan, 1987, pp. 93-111];

issue. This opens up the possibility of responding to the argument from atomism in a new way. In what follows I shall focus on one corollary of this novel reply to the argument from atomism, viz., that the experience of secondary qualities involves an illusion.

DIRECT REDUCTIVE COLOR REALISM

But before zeroing in on the problematical role of illusion I shall try to place this new line of argument into a broader context. Focussing 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 phenomenal red I currently experience 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 unaided by any intermediary entities, like, e.g., sense data. Colors are apprehended non-inferentially, unmediated by any kind of inference from sense data or other such causal middle-men as are posed by 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 qualities. What we are dealing with, then, is a form of direct, reductive color realism, or *reductive realism* for short. Armstrong, who

is the most tireless defender of reductive realism, has characterized it succinctly:
"[S]econdary qualities are primary qualities not apprehended as such."
 [Armstrong and Malcolm, 1984, p. 185, italics added]

The reductive 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 qualities 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 and Malcolm, 1984, p. 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 problematical. 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. Giving up the irreducibility of phenomenal properties is an ingenious move on the part of the reductive realist. It enables her to achieve two goals at once. First, it makes it

possible for her to acknowledge the power of the argument from atomism *in the context in which it was proposed*. This is a welcome consequence. For if you claim the untenability of a 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 validity given the historically prominent irreducibility assumption, the reductive realist does explain the argument's robust appeal. Second, in giving up the irreducibility assumption the reductive 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. For if secondary properties are primary properties, then matter can have secondary properties as well.

Now that we have a better understanding of reductive realism and the importance it attaches to the reducibility of phenomenal qualities, I can finally state the problematical corollary of reductive realism. Armstrong admits that there is a price to pay for the reduction of secondary qualities: the reductive realist has to admit that "*sensory illusion is a much more widespread affair than we ordinarily assume.*" [Armstrong, 1961, p. 164] Writing sixteen years later he acknowledges that, on his theory, the "perception of secondary qualities involves an element of illusion, " [Armstrong, 1977, p. 31] and his most recent way of putting it is that his view has the consequence that we suffer from "the illusion of concrete secondary quality." [Armstrong and Malcolm, 1984, p. 180] There can be no doubt that the appeal to illusion is a crucial and enduring trait of Armstrong's reductive color realism. Therefore reductive 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.⁶⁸

THE ILLUSION OF CONCRETE SECONDARY QUALITY: ARMSTRONG'S ACCOUNT

We can begin by excluding one obvious interpretation of the illusion in question as unacceptable for the relocationist. On this interpretation the illusion consists in our failure to locate the experienced phenomenal qualities correctly. We take them to qualify the external object of perception. But really they are properties of images, sense data, or some other mental category. I think that Eddington's claim that the ordinary table of common sense is an illusion is most naturally interpreted along these lines. If this were the correct interpretation of the kind of illusion involved in reductive realism, reductive realism would clash with the

⁶⁸ Armstrong is not alone in his appeal to illusion when it comes to explaining how the phenomenal qualities of experience fit into the physical world. Sellars takes this possibility very seriously (but ultimately rejects it) when he considers how the sense qualities fit into the neurophysiological image. According to him, the following is a possibility: "the neurophysiological image is complete and the ultimate homogeneity of the sense qualities (and, hence, the sense qualities, themselves) is *mere appearance* in the very radical sense of not existing in the spatiotemporal world at all." [Sellars, 1962, p. 36]

Nor is the appeal to illusion limited to those who are interested in defending a form of reductive color realism. If anything the importance of the role of illusion is even greater in other approaches to secondary qualities. Take, for example, Lycan's view that qualia are borne by nonexistent sense-data. In reply to the objection that one can *see* the actual green after image that one is having he asks us to consider the possibility that one is suffering from an illusion. Lycan asks: "how might the believer defend this last claim against the competing hypothesis that it merely *looks to him as though* there were something green before him, when in reality there is not? [Lycan, 1987, p. 88] Eliminativists constitute another group who need to appeal to illusion in order to secure any plausibility for their claim. Hardin, for example, puts it as bluntly as you could wish for: "Colored Objects are illusions." [Hardin, 1988, p. 111] Nor has Dennett shied away from appealing to illusion when it comes to eliminating (mental) qualia: "Philosophers have adopted various names for the things in the beholder (or properties of the beholder) that have been supposed to provide a safe home for the colors and the rest of the properties that have been banished from the "external" world by the triumphs of physics: "raw feels," "sensa," "phenomenal qualities," "intrinsic properties of conscious experiences," "the qualitative content of mental states," and, of course, "qualia," the term I will use. ...I seemed to be denying that there are *any* such properties. But...I agree wholeheartedly that there seem to be qualia." [Dennett, 1991, p. 372]

relocation thesis. Since reductive realism is supposed to support the relocation thesis, we shall have to search for a different interpretation for the kind of illusion involved.

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 and Malcolm, 1984, p. 178]

Our experiential grasp of the secondary properties is misleading: we experience simplicity and continuity where there is structural complexity. The phenomenal quality we experience appears to be remarkably unlike what it is supposed to be on the reductive realist view. Take, for example, the secondary property of phenomenal greenness: the property 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:

[W]hen 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 and Malcolm, 1984, p. 179]

So far reductive realism has fared badly. The talk of illusion has proved pernicious. Instead of the intended reduction of ontological baggage reductive realism seems to have lead to an increase in the number of properties. In addition to primary and secondary properties the reductive realist needs to

countenance properties that might be called *primary property counterparts of secondary properties*: 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 reductive realism has not achieved its goal of increasing ontological economy.

In order to make some progress the reductive realist must find a way to identify secondary properties with their primary property counterparts. Armstrong attempts to facilitate this identification by radically overhauling our notion of secondary properties. By employing the "topic-neutral maneuvers" 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 and Malcolm, 1984, p. 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 properties 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, p. 275]

Armstrong is well aware that this account of secondary qualities faces “an enormous phenomenological difficulty.” [Armstrong and Malcolm, 1984, p. 179] For in experience we seem to become fully acquainted with the striking intrinsic natures of these phenomenological properties. The outrageous implausibility of reductive 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 two considerations to unveil the illusion of concrete secondary quality. First, we suffer from an illusion that is similar to the headless woman illusion. Our failure to detect any complexity in the phenomenally given secondary properties creates the impression that we do 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 and in particular the colors. Unaware of the real primary quality basis of these resemblances and differences between colors we erroneously take ourselves to have an awareness of the intrinsic natures of the 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 and Malcolm, 1984, pp. 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 qualities. Our mistaken impression that we are very well acquainted with the inner 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. And 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.

THE ILLUSION OF CONCRETE SECONDARY QUALITY: SOME PROBLEMS

Armstrong's defence of reductive realism is impressive. Nevertheless I find one of its central features puzzling and obscure. This feature is the reductive realist's appeal to illusion. Is the appeal to illusion an indispensable feature of reductive realism? I believe that this is so. The appeal to illusion is what saves *reductive* realism from becoming an eliminativist theory of qualia.

When confronted with the patent difficulty that lies in identifying color qualia with primary quality complexes, Armstrong makes two coordinated moves: the topic-neutral maneuver and the appeal to illusion. Both moves are indispensable components of reductive 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 reductive realism at this point we would be dealing with an eliminativist theory of qualia: qualia as we experience them continuously would have been lost.

This is where the second move—the appeal to illusion—comes into play. Its purpose is to restore the phenomenological adequacy of the reductive realist account. The phenomenal aspect of phenomenal properties (that was banished by the topic-neutral maneuver) reenters the picture in the guise of an illusory component of our experience of phenomenal properties. Thus everybody is supposed to be happy. The qualophile has been provided with his warm phenomenal glow upon which he insists. And the scientifically minded reductionist has been furnished with well behaved phenomenal properties that do not buck identification with scientifically respectable properties.

This, then, is the role of the appeal to illusion in reductive realism: the appeal to illusion prevents the theory from collapsing into a form of eliminativism. 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 reductive realist project.

In what follows I argue that Armstrong's appeal to illusion is doubly problematical. First, and less importantly, I argue that Armstrong's account of the origin of the illusion of concrete secondary quality is quite unconvincing. But this does not really go to the heart of the matter. A better story about the genesis of the illusion may well be available. Second, and more importantly, I argue that a satisfactory interpretation of the illusory element is not available. It appears that the reductive realist is forced to admit that our experience of phenomenal color is wholly illusory. As a consequence the reductive realist's account of veridical experience of phenomenal color inherits all the problems that beset his account of nonveridical experience. These difficulties are compounded by additional problems raised by the problems created by the argument from atomism. Hence the attempt to save reductive color realism by demonstrating its power in its chosen field of veridical experience must be regarded as a complete failure.

ON THE ORIGIN OF THE ILLUSION OF CONCRETE SECONDARY QUALITY

I want to emphasize just how speculative and astonishing the last step of Armstrong's defence of reductive realism is. Armstrong's account of the origin of the illusion of concrete secondary quality is astonishing 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 has 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 these perceived

resemblances 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 served no purpose at all. The illusion does not advance our understanding of the problematical fact at all.

There are other philosophical “feigning” theories that fare much better. Hume, for example, proposes 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 with an unchanging self? Hume holds that we can pull off this feat by feigning acquaintance with a self that introduces the needed unity into the diversity. He says: “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 perceptions 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 (1739), p. 254] Hume’s story makes sense: the feigning he ascribes to us helps to explain how we manage to believe in personal identity; the self is the right kind of thing to give us personal identity. Armstrong’s story, on the other hand, lacks this virtue. The feigned

acquaintance with intrinsic properties of colors does not help explain the perceived similarities between them. For this reason Armstrong's speculative account of the origin of our belief that we are directly acquainted with the intrinsic natures of secondary properties hardly seems plausible.

ON THE NATURE OF THE ILLUSION OF CONCRETE SECONDARY QUALITY

The appeal to illusion plays a crucial role in Armstrong's version of reductive realism. One cannot hope to understand the theory without a clear grasp of the nature of the illusion involved. Unfortunately I find myself incapable of determining the precise character of the illusion of concrete secondary quality. As best I can determine the appeal to illusion remains unintelligible within the reductive realist framework. I argue that the appeal to illusion burdens the reductive realist with all of the problems that we ran into in our attempt to make sense of nonveridical 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? My search for an answer to this question shall be guided by the maxim that I should pay especial attention to the phenomenal facts involved. The phenomenal red is, after all, the problematical 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 reductive realism.

This maxim cannot be pushed aside as something that is gratuitously grafted upon reductive realism. A condition on the adequacy of the reductive realist analysis of phenomenal red is that, after all is done and said, we are still

left with *phenomenal* red. For reductive realism is not supposed to be an eliminativist or disappearance theory of phenomenal properties. As a reductive theory reductive realism tells us that phenomenal properties do really exist and it tells us what their true nature is. Thus, whatever the final reductive realist analysis of phenomenal red is, the analysans with which the reductive realist comes up must be classifiable as phenomenal red in just the same way in which the analysandum with which he started out was a phenomenal red. For if we start out with phenomenal red and end up with something that is not phenomenal red then we are no longer dealing with a reductive analysis. 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 reductive realist is given the magical power to make it so that the tomato I see has the property with which he identifies phenomenal red. Then a minimal condition on the correctness of the reductive 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: the reductive realist analysis of phenomenal red has to save the phenomena.

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 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 qualities *and nothing else*? What would the surface of the tomato look like if it were experienced as having “bare” primary qualities? 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 quality abstracted from all other sensible qualities “implies a repugnancy in its conception.” [Berkeley, 1985 (1713), p. 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 phenomenon 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 as illusory, reductive realism turns into a 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 reductive realist analysis of phenomenal redness then I should regard reductive realism about color as a complete failure.

SECOND INTERPRETATION: PARTIAL ILLUSION

Here is a different model with which to interpret the notion of “disillusioned” experience. Consider the Benham disk.

The Benham disk. If the disk is rotated at a rate between 5 and 10 times per second and viewed under a bright tungsten light, desaturated bands of various colors will be clearly seen. Their sequence will depend on the speed and direction of rotation of the disk. [Hardin, 1988, p. 73]

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 a black and white pattern. If we could “stop the tomato” in a manner analogous to which we can stop the rotating Benham disk, then we might be able

to apprehend the complex of achromatic properties, that is its phenomenal redness, *as such*, i.e., as a black and white pattern.

What does most of the work in this analogy is the switch from the opposition between primary and secondary properties 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 properties is equalled 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 problematical 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 reductive color 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 qualities. 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 reductive realism postulates in the ordinary experience of phenomenal color we shall repeat the experiment of imagining what a "disillusioned" experience would be like. If I managed to "stop the tomato" then

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 reductive 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.

THIRD INTERPRETATION: MINIMAL ILLUSION

The first two attempts to understand the illusory component of reductive realism went wrong by making the illusory component too pervasive. Not only some aspect of phenomenal redness turned out to be illusory; phenomenal 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 Armstrong's remarks on this issue. 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 and Malcolm, 1984, p. 174]⁶⁹ 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, p. 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 and Malcolm, 1984, p. 179]

Let us focus on the second kind of illusion, the illusion that "phenomenologically, the secondary qualities lack structure." [Armstrong and Malcolm, 1984, p. 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 lead 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 of the painting or mosaic: the homogeneously colored expanse dissolves into a multitude of colored (sometimes multicolored) specks. Thus the illusion of structurelessness

⁶⁹ See also [Armstrong, 1977, p. 30]

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 reductive color 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 reductive color 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 doubtful 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 reductive realist perspective because the analysis 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 qualities or in terms of

complex achromatic qualities, we run into the problems that were surveyed above.

REDUCTIVE COLOR REALISM AND ILLUSORY COLORS

I find myself unable to provide a satisfactory interpretation of the nature of the illusion that, according to reductive color realism, is involved in every experience of any phenomenal color property.⁷⁰ 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 (whatever other aspects there might be to phenomenal red) into the realm of illusion. While this move may smooth the path to a reductive account of phenomenal red, it does leave us with the problem of accounting for these illusory colors. To say that they are unreal because illusory and are, therefore, not in need of further explanation, will not do. For 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 intentional objects seems inevitable. If the real tomato is not allowed to bear the

⁷⁰ 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.

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 reductive 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 veridical experiences. But the close scrutiny of the case of veridical experiences of phenomenal color has proven this impression wrong. The reductive realist is forced to use the very same problematical notions in his analysis of veridical and nonveridical experiences of phenomenal colors. Thus reductive color 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 reductive realist interpretation of the relocation thesis.

It is worth pointing out that the second encounter with merely intentional bearers of phenomenal properties promises to be more problematical than the first. For this time around we have to consider the further question whether merely intentional physical objects are capable of supporting seemingly simple and homogeneous phenomenal properties. That is, the problems created by the

argument from atomism forced the reductive realist to attribute phenomenal properties to merely intentional objects. But it would seem that the argument from atomism applies all over again at this new level: merely intentional physical objects are no more hospitable to secondary qualities than real physical objects. An unedifying regress looms ahead.⁷¹

S U M M A R Y

In this Chapter we took on the first argument for the relocation thesis: the qualia must be located on the mind's object because the mind itself cannot house qualia. In reply, two traditional objections to the relocation move were raised. First, the argument from illusion: what do we do with nonveridical experience where there is no external object to bear the extruded qualia? Second, the argument from atomism: qualia are not the kind of property a physical object could bear.

In response to the problems raised by nonveridical experience the relocationist claimed that (at least in these problematical 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 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

⁷¹ See [Lycan, 1987, p. 90-91] for a discussion of this issue. Interestingly he thinks that Armstrong's "Direct Realism" offers a solution for 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.

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 ends cannot be served by an intentional object that is a mere possible.

It seems, then, that nonveridical experience poses a formidable problem to the relocation thesis. 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 thesis in the domain of veridical experience. I focus on Armstrong's reductive realist attempt to deal with the argument from atomism. Armstrong grants the atomist his conception of matter, but revises our notion of a secondary quality. By claiming that secondary qualities really are complexes of primary qualities, Armstrong can embrace the atomistic picture of matter and claim that the phenomenal properties are properties of matter. But to save the phenomena the reductive 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. The element of illusion in reductive color realism is much greater than would appear on first sight. Phenomenal color turns out to be entirely illusory. Thus it seems that the relocationist who attempts a reductive realist explanation of veridical color experience can only succeed if he has an adequate account of nonveridical or illusory color experience. And 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 de-mented 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 follows 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 shall propose a way of resolving this impasse. I shall 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 6 CONSCIOUSNESS: THE HAVING OF QUALIA EXPLAINED

LEARNING FROM MISTAKES

To be conscious is to have qualia. How much progress have we made in elucidating this somewhat cryptic slogan? Lets begin by taking stock.

First, What have we learned about the having of qualia? I have argued that to have qualia is *not* a matter of knowing them infallibly, *not* a matter of knowing them fallibly, and *not* a matter of believing invariably that you have them. I went on to argue that this bears negatively on the thesis that the having of qualia is a matter of introspecting them. These were the conclusions reached in Chapter 2. In Chapter 3 I argued that consideration of Oscar teaches us that having qualia (in the way that makes for consciousness) is *not* a matter of introspecting them. For we should think of Oscar as a machine that has the introspection but lacks the consciousness. In Chapter 4 I introduced a further consideration concerning the having-question. I argued that if you make qualia into nonmental properties of external objects, then the question of what it is to have them becomes intractable. I failed to show that an account of the having of de-mented qualia could not be had. But I had to admit that I do *not* know what such an account might possibly look like.

Second, What have we learned about the nature of qualia? I started out with the traditional assumption that qualia are mental properties and supported this assumption with the first appeal to phenomenology. But consideration of the second appeal to phenomenology undermined the phenomenological evidence for

the mentality of qualia. The following consideration seemed to clench the case against the mentality of qualia: qualia cannot be mental because there are no mental items fit to function as bearers of qualia. So qualia are *not* mental properties. This lead us to consider the possibility that qualia are nonmental properties of nonmental, external, objects. But I argued that qualia cannot be nonmental properties because there are no physical items fit to function as bearers of qualia. So qualia are *not* physical 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.⁷² 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 are strong enough to suggest a direction for further research. To me the negative results of the preceding chapters suggest the following guidelines for the upcoming investigation. We want unrelocated qualia. For I do not have a story to tell about what it is to have relocated qualia.

⁷² Actually things are really worse on the question of the nature of qualia. It seems that there cannot be any qualitative phenomena at all since nothing in the world can exemplify qualia. So perhaps we have learned one "positive" thing about qualia, viz., that there are none. In what follows I hope to show how to avoid this conclusion.

We want bearer-less qualia. For there are no qualia-bearers, neither in the mind nor in the physical world, be they actual or possible. We want self-presenting qualia. For neither knowing (fallibly or infallibly), nor believing, nor introspecting lets us *have* our qualia in the required way. Since I do not know which other activity of the mind might serve the purpose of presenting us with our qualia in a way that makes us conscious, I opt for the notion of self-presenting qualia. What I shall propose, then, in this Chapter is an account of consciousness in terms of unrelocated, unborne qualia the having of which consists in their presenting themselves to us.⁷³

ISOLATING TWO PERNICIOUS ASSUMPTIONS

We want to make sure that the positive account of consciousness avoids the mistakes of its rivals. We will, therefore, be well advised to let the construction of the positive account of consciousness be informed by what we can learn from the failure of its rivals. But since one can learn from mistakes only after one has clearly grasped their nature, I propose to begin this inquiry with an attempt to isolate the roots of the problem from which the rival accounts suffered.

The problems we encountered were of two kinds. First there is the problem of the intractable nature of the qualia-bearing objects of experience. Second there is the problem of the nature of the mental act that makes us have

⁷³ 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.

our qualia. These two problems 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? Are they actual or merely possible objects? Which of these candidate qualia-bearers can actually exemplify phenomenal properties? Which relation must a qualia-bearing object have 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, and introspection have been excluded as possible vehicles for the having of qualia? Perhaps these questions can be answered satisfactorily, but if so, I do not know how. Therefore I must attempt to construct the positive account in a way that keeps all 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 arises because of two unquestioned (and often tacit) assumptions. The first assumption 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 shall 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 shall call this the *assumption of the substance/attribute structure of qualified objects* ("substance/attribute assumption" for short). If we can come up with a conception of qualified objects that does not require an underlying substance or substratum then the problem of the intractable nature of qualia-bearing objects experience vanishes. And if we can come up with a nonrelational analysis of experience with which to replace the act/object model

then the problem of the nature of the mental act that makes us have qualia vanishes too.

Our goal must be, then, to present an account of consciousness that avoids both of these assumptions. I shall discuss two theories of consciousness that get by without making these assumptions. These two theories are the adverbialist theory of consciousness and the neutral monist theory of consciousness.

Adverbialism gives up on the act/object structure of experience by, in effect, making the qualified object into a modification of the subject. Adverbialism is an elegant and radical way of dispensing with the two problematical assumptions. The act/object assumption and the substance/attribute assumption are assumptions one makes 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. The adverbialist does not countenance qualia understood as phenomenal properties. Thus the questions What supports qualia? and What is it to have qualia? don't even so much as arise.⁷⁴

⁷⁴ A sympathetic discussion of the adverbialist theory of consciousness may seem out of place in a chapter that tries to develop a theory of consciousness that explains consciousness as the having of qualia. For doesn't adverbialism deny the existence of qualia? And wouldn't the truth of adverbialism, therefore, show that consciousness is not a matter of having qualia? I agree with this worry. Were I to convince myself that adverbialism is true I would give up the claim that consciousness consists in the having of qualia. 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 would be dealing with elimination, not reduction. Well, if this is so, why then discuss adverbialism in this chapter? My reason for doing so is that the way I set up the problem cries out for an adverbialist answer. For adverbialism is surely the first theory that comes to mind when we are told to develop a theory of consciousness that avoids that act/object assumption and the substance/attribute assumption.

The neutral monist theory of consciousness is not nearly as radical as adverbialism. That is, the problematical kinds of questions the mistaken answers to which are driven by the act/object assumption and by the substance/attribute assumption do arise within the neutral monist framework. The neutral monist does accept the existence of qualia understood as phenomenal properties. But by invoking the bundle theory he can avail himself of the conception of a qualified object that does not commit us to some qualia-bearer or qualia-bearing substance.⁷⁵ And by collapsing act and object into one he arrives at a notion of a self-presenting quale that can be viewed as an answer to the having problem.

THE ADVERBIALIST THEORY OF CONSCIOUSNESS

Adverbialism is most easily (but not exclusively) 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:

[W]hat 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

⁷⁵ 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 any 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 had is, then, one of the major liabilities that my account of consciousness is burdened with.

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, p. 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 thesis 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.

I now turn to the critical consideration of the adverbial theory of experience. I shall argue that adverbialism, when spelled out more fully, collapses into a form of the act-object model of sensation. Therefore it does not keep the promise of affording a genuine alternative to a relational analysis of experience. To avoid the fantastic complication of the adverbs that characterize the kinds of sensing occurring in situations in which, as we would ordinarily say,

we experience a complex object, I shall focus on the simplest possible case. Let us look at a case which is fully described by saying that I sense something red, i.e., a case in which I sense redly. How is the sentence “I sense redly” to be understood? More particularly, how does the word “redly” function in this sentence?

We can make a good beginning by looking at the critical questions that Chisholm asks of the later, reistic, Brentano’s account of intentional objects. It seems that, after the necessary changes have been made, these same questions can be asked of the adverbial analysis of experience. The prospect of using Chisholm’s critique of Brentano’s reistic account of intentional objects to criticize the adverbial theory of experience is particularly enticing because Chisholm is one of adverbialism’s leading advocates. 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, p. 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 another place [Chisholm, 1982, p. 63] Chisholm uses these same considerations to support a Meinongian construal of intentional objects. What I want to do is to ask the same kinds of questions that Chisholm raised about “John is thinking about a unicorn” about the sentence “John is sensing something red.” 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 word “red” as it occurs in “John sees 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. [6] We are saying—again obviously—that the *object* of John’s sensing is red. But, the adverbialist tells us, [7] 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 [6]. 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 the thing that is John’s intentional object is obvious to all but those who have already embraced adverbialism. [6] only states what the adverbialist knows. It causes him no embarrassment.

I turn to sentences [1] and [7]. 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 [7] are true of his position. But, again, these two claims need not cause the adverbialist any embarrassment.

[2] and [3] are unproblematical. 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 describe 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, bordering on the senseless. To say that John's sensing is red is no saner 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 we can draw the following conditional conclusion: If Chisholm's case against Brentano's reistic interpretation of intentional objects is convincing then our case against the adverbialist is strong too.

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:

[F]requently 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, p. 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, p. 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 kind *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 problematical sentence [4] the adverbialist can avoid the unwelcome consequences I wanted to pin upon him. It would seem, then, that the conditional conclusion that was reached above—if Chisholm’s case against Brentano’s reistic interpretation of intentional objects is any good then our case against the adverbialist is strong too—has to be given up. For the adverbialist has answered the Chisholmian questions.

But the adverbialist gambit leaves me dissatisfied. I can express my worry in the form of two questions. First, 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 talk about certain species or manners of sensing that does not boil down to understanding it as talk about the sensing of things that are certain ways, 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, p. 96] But this analogy does not help to improve our understanding of the problematical “redly” in “senses redly.” I can do no better than endorse what Panayot 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, p. 265]

Thus we are faced 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, p. 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 lead 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 problematical 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.

The second question I want to raise can be put in a number of ways: 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 hasn’t the adverbialist already explained that the experience occurs in a red manner; that the experience belongs to a more determinate species of the genus *sensing*, viz., the determinate species *sensing redly*? Of course the adverbialist has said that. But may we not ask the adverbialist what makes my experience be of the particular species *red*? May we not ask on strength of what my experience occurs 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 a part of the correct answer. For 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 the above 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 sensing states does not solve the problem the adverbialist set out to dissolve. The phenomenal properties are back and cry out for objects of some kind to support them. And to get rid of these objects was what motivated the adverbial analysis of experience in the first place. Thus the adverbialist must reject the above proposal, viz., to explain the notion of species of experience by appeal to the corresponding phenomenal properties, as simply being silly.

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 our question, viz., In virtue of which of its intrinsic features does my sensing event belong 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.

But 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. Neither 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 the adverbial 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 far short of a proof that adverbialism is false, it does direct us to investigate competing nonrelational analyses of experience.

THE NEUTRAL MONIST THEORY OF CONSCIOUSNESS

Adverbialism attempts to overcome the act-object model of experience by, in effect, turning the object (with its phenomenal properties) into a modification of

the subject. What used to be the object is now the manner in which the act occurs. Adverbialism's merger of act and object proceeds at the expense of the object: it is absorbed into the act, as it were. The resulting monism leans precariously towards the subject's pole of the act object distinction. As is indicated by its name, neutral monism attempts a less lopsided unification of act and object. By identifying act and object neutral monism ends up with a notion of experience that is as much subjective as it is objective. The neutral monists preferred to express this fact by saying that experience or sensation is neutral as between the act and the object or, more grandly and less accurately, by saying that experience is neutral as between mind and matter.⁷⁶

THE ROUTE TO SELF-PRESENTATION : IDENTIFYING ACT AND OBJECT

Why would anyone perform this rather unlikely identification of the act and the object of a sensation? What moved the Neutral Monists was an unwillingness to countenance unobservables paired with a great respect for the keen blade of Ockham's razor. I shall support this claim by looking at Bertrand Russell's reasons for endorsing neutral monism. Other champions of neutral monism like,

⁷⁶ I use the label "neutral monism" to stand for a particular form of nonrelational account of experience or sensation. I should point out that this is a rather idiosyncratic (though not entirely unjustifiable) use of the term. Historically the Neutral Monists like, for example, William James, Ernst Mach, and Bertrand Russell took their theory to have a vastly greater scope than this. As they saw it, Neutral Monism was a comprehensive 'metaphysical' system on a par with such metaphysical doctrines as materialism, idealism, or dualism. What I call neutral monism here is a much more limited doctrine. It is the account of experience you get by identifying the two poles of the act/object model of experience. I do not (yet) mean to endorse a general metaphysics that constructs the world from neutral elements. In order to avoid unnecessary confusion I shall use the following convention from now on. I write "neutral monism" when I want to talk about the limited doctrine concerning the nonrelational nature of experience; I write "Neutral Monism" when I want to talk about the grand metaphysics.

for example, William James and Ernst Mach, were motivated by similar considerations. Russell reasons as follows: the subject or act cannot be observed; therefore the subject must be given up; therefore the relational theory of sensations must be given up. On the unobservability of the act he has this to say:

[T]he *act* seems unnecessary and fictitious. ...Empirically, I cannot discover anything corresponding to the supposed act; and theoretically I cannot see that it is indispensable. We say: "*I think so-and-so*," and this word "*I*" suggests that thinking is the act of a person. [The] "*act*" is the ghost of the subject, or what once was the full-blooded soul. [Russell, 1978 (1921), pp. 17-18]

Later on he bases the identification of act and object on the same kind of consideration:

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 perfectly 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 (1921), pp. 141-142]⁷⁷

⁷⁷ See also [Russell, 1919, pp. 305-306] and especially [Russell, 1975 (1959), chapter 12 "Consciousness and Experience"].

This consideration proves to Russell that “the patch of color and our sensation in seeing it are identical.” [Russell, 1978 (1921), p. 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?⁷⁸ 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 the abolishment of the act. 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 object sensed—the sense-datum—the sense-datum is “presentationally inert.” It does not force itself upon us. It is we who make it

⁷⁸ 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 (1921), p. 22]

our own by reaching out towards it in 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 are otherwise inevitable, we must find a new name of the successor concept of the sense-datum. Appropriating a term that Russell introduces in a slightly different context, we can christen these successor concepts "percepts." Adopting this usage I can put the point as follows: upon discovering that the subject does not exist Russell is lead to identify the sensing with the sense-datum because he sees that the resulting percept has to play to role of both of the original terms of the sensing relation. The resulting notion of a percept is the notion of a *self-presenting state or object*.

APPLYING NEUTRAL MONISM: THREE QUESTIONS

While the preceding considerations do, hopefully, throw some light on the neutral monist theory of experience, they failed to show how this theory bears on the problems generated by phenomenal properties. That is, what we have achieved, so far, is, at best, a rudimentary understanding of the theory; not an understanding of how the theory is supposed to solve our problems. I want to raise three questions that a neutral monist theory of experience will have to deal with. First, what is it for us to have a phenomenally propertied state, i.e., a

percept? Second, does this theory provide phenomenal properties with a home? Third, is the theory's account of experience phenomenally adequate?

WHAT IS IT TO HAVE A PERCEPT? THE BUNDLE THEORY OF THE SELF

What is it to have a phenomenal state? 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 need to be said about it.

I think that the reason that this answer does not satisfy is that the notion of self-presentation exudes the foul smelling aura that is characteristic of the miraculous. I smell 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 any less miraculous than the notion of a self-presenting state? I admit that we are much more used to attribute such presentational powers to mental acts. But how the act is supposed to accomplish this remarkable feat of presenting us with a phenomenal particular remains in the dark. 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 this case 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 monist 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 monist 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 just bad grammar.

This is a fair objection; but one that the neutral monist can answer, if not without difficulty. I shall only attempt the barest sketch of an answer that the neutral monist’s might give. The centerpiece of this answer will be a bundle theory of the self. The neutral monist will deny that he employs the absurd notion of the presentation of something to nothing. 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. The neutral monist no longer conceives of presentation as something akin to the confrontation between two individuals; rather he sees it as matter of membership in a unified group of percepts.

This, then, is what the neutral monist will say about the questions what it is for a percept to be presented 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 percept-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 any of the other percepts 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 which is the bundle.

A HOME FOR QUALIA? THE BUNDLE THEORY OF THE PHENOMENAL OBJECTS AND UNBORNE QUALIA

The attempt to avoid the difficulties encountered by the act/object analysis of experience was what motivated the investigation of nonrelational theories of experience. One of the main difficulties of the act/object model was the intractable nature of the qualia-bearing objects. We found that none of the candidate objects could provide a satisfactory solution to both the exemplification problem and the causation problem. 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 to 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 will not get solved by discovering some special kind of qualia-bearing object. The all-important first step towards the 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 intentional 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. And at the same time he provides a new and unexpected function for qualia to play. By turning qualia into adverbial modifications of a subject the act (by means of which the subject appropriates her qualia) is rendered superfluous. In this respect adverbialism must be counted a success.

How does neutral monism fare on this account? In what manner do the phenomenal properties enter into the percept that results from fusing the sensing with the sense-datum? In which sense of “in” are the phenomenal properties in the percept? So far the ontology of percepts has not been addressed. What ontological category do percepts belong to? If the neutral monist answers that percepts are objects of some kind that have the phenomenal properties he faces the same difficulties with which the friend of intentional objects had to contend. I have repeatedly emphasized the problems of the view that qualia can be taken as properties of some kind of object: the right kind of qualia bearing objects do not exist. If the final word on neutral monism is that percepts are substances the attributes of which are phenomenal properties, then neutral monism is in trouble.

But the neutral monist does not get impaled on this problem. Very much in keeping with his radical empiricism and his allegiance to the principle of ontological parsimony he manifests a marked abhorrence of substances. Therefore 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 neutral monist will say that a percept just is the bundle appropriately united phenomenal properties. Let us again turn to Russell in order to see how he turns the corner to the bundle theory of mental particulars.

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 (1940), p. 98]

True to this self-imposed directive Russell proposes to trade substantial things for things that are constituted solely by their properties:

One is tempted to regard “this is red” as a subject-predicate proposition; but if one does so, 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. Such a view is open to all the familiar objections to the notion of substance. ...I wish to suggest 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 (1940), p. 97]

When Russell discusses the problem of how to introduce the notion of “place” into this construction he describes a situation in which one experiences two red patches. For our purposes the passage is interesting because it makes clear that he means to be talking of mental “things” and moreover the passage also gives us an indication of the kinds of qualities from which Russell intends to construct percepts.

Suppose we see simultaneously two patches of a given shade of color C ; let the angular coordinates of the one patch in visual space be θ, ϕ , and those of the other θ', ϕ' . Then we are to say that C is at (θ, ϕ) and also at (θ', ϕ') . The angular coordinates of an object in the visual field may be

regarded as qualities. Thus (C, θ, ϕ) is one bundle of qualities, and (C, θ', ϕ') is another. If we define a “thing” as the bundle of qualities (C, θ, ϕ) , then we may say that this “thing” is at the place (θ, ϕ) , and it is analytic that it is not at the place (θ', ϕ') . [Russell, 1980 (1940), p. 99]

Summing up the main features of this construction A.J. Ayer points out the following features:

[Russell] argues that it is possible, at least at the lowest logical level, to dispense with substances. Sensory properties do not need a *substratum* to support them, and the work of holding them together can be done, he now thinks, by a relation of compresence. ...The complexes of qualities which names denote include spatial and temporal qualia, conceived as constituents of private sense-fields. [Ayer, 1971, pp. 45-46]

What we have here is yet another theory about what it is for a phenomenal property to inhere in what it qualifies. We already looked at two other theories of what it is for something to have 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 lead 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 lead 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. The adverbial theory fails on this account: the role it assigns to phenomenal properties remains obscure. What is good about the adverbial theory is that it gets rid of the peculiar intentional objects that supposedly are needed to support the phenomenal properties. The substance/attribute theory fails on this account: its need for substrata, for supports of phenomenal properties, causes numerous problems. 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.⁷⁹

But wait! Property instantiations without property instantiators do not make any 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

⁷⁹ 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 not an adherent of the relational theory 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 untempered by the adoption of the bundle theory.

“unsubstantiated” property exemplification. How should one reply to this challenge? I fear that the best I can do is to admit the obscurity but, at the same time, to point out a parallel obscurity in my opponent’s position. A property instantiation unsupported by a substance is a queer thing. And 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 serve to shut him up by pointing out the generality of the problem he has raised. The nature of property instantiation is one that everybody (who is not a nominalist) has to deal with. On this count, at least, the neutral monist is no worse off than the rest.⁸⁰

How, then, does the neutral monist propose to house the qualia? On one reading the neutral monist makes qualia into homeless objects. After all, he robs them of a substance in which to inhere in. But this reading is needlessly misleading. A more illuminating way to put the point is this. The phenomenal red that I experience upon 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

⁸⁰ 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 problematical 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 is committed to the view that the bundles are bundles of properties traditionally understood.

there is a sense in which the neutral monist can acknowledge that the phenomenal red of the tomato is borne by an object. The object is, however, not a problematical 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 neutral monist proposes to house qualia without getting stuck with a problematical qualia-bearer.

THE PHENOMENAL ADEQUACY OF NEUTRAL MONISM

Phenomenological adequacy is the first requirement that any theory of experience must satisfy. If “trying on” (Searle’s term) the theory does not yield an experience that is like what we usually experience, the theory must be given up. I have criticized Armstrong’s version of reductive color realism on these grounds. And I expressed my worry that adverbialism is inadequate to the task of preserving the phenomenal features of experience.

As far as I can tell, the neutral monist theory of experience passes this test unscathed. The neutral monist theory of experience 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; and it does not reconstruct the logic of phenomenal property terms in a way that lets one doubt whether undergoing such a logically revamped experience would be anything like undergoing a regular experience. Phenomenologically speaking, neutral monism leaves everything the way it is. The changes are in the ontology; the

phenomenology stays the same. Neutral monism tells a revisionary story about what you experience; but how you experience it remains unchanged.

FEELING QUALIA: THE NONCOGNITIVE NATURE OF CONSCIOUSNESS

I have argued that the combination of neutral monism and the bundle theory provides a phenomenally adequate account of consciousness. It does so by telling us what qualia are and what it is to have them. I now want to discuss a consequence of the neutral monist's account of what it is to have qualia. I shall argue that the having of qualia is not a matter of knowing anything in a special way, not a matter of knowing anything in a fallible way, not a matter of believing anything, and not a matter of seeing anything with an inner eye. To have qualia is not to cognize them in any way. I shall express this by saying that we *feel* our qualia. I shall end by discussing a challenge to the view that consciousness is a noncognitive state. I conclude that the dispute about the (non)cognitive nature of consciousness is primarily a dispute about words that leaves the substance of the neutral monist account of having qualia intact.

Percepts may fittingly be characterized by the claim that "Berkeley's axiom that *esse est percipi* is true of these states and so is Sextus Empiricus's concept of an object that it apprehended *through itself*." [Hess, 1988, p. 5] In the percept there is no duality between its existing and its being experienced, between one's having a percept and one's being conscious of it. Two important consequences flow from this feature of percepts. First, it is in virtue of this feature of percepts that consciousness without introspection is possible. Percepts need no introduction: they announce themselves and need not be introspected to

make their presence felt.⁸¹ Perhaps creatures endowed with introspective capacities are capable of introspecting their percepts. But if they are it is not in virtue of this capacity that they enjoy qualitative consciousness. Perhaps this reflective form of self-awareness should be called consciousness of consciousness—the cognition that one is conscious. But if consciousness is what we are interested in, then we need pay no attention to such higher level processes.

The second consequence of the fact that percepts are act and object melted into one may, perhaps, be less welcome. It is this: our way of having percepts must be characterized as *noncognitive*. The having of a percept is not a matter of knowing something in an epistemically special way; not a matter of knowing something in an ordinary fallible way; not even a matter of believing or accepting something. How, then, can this “having” be characterized in a more positive vein? Perhaps we should stick with the minimalist answer that Moritz Schlick gave in the Chapter he devotes to attacking the “so-called internal perception:” “The [percept] simply *is* there, and this disposes of the whole matter.” [Schlick, 1985 (1925), p. 153] This is the least misleading way to talk about the matter. A large number of terms have been used in the attempt to provide a more full-blooded rendition of the pale having-idiom: “apprehend,” “experience,” “sense,” and “feel” come to mind. All of these terms are

⁸¹ In criticizing Locke’s claim that the mind has the capacity to perceive its own ideas Schlick makes just this transition from the *esse est percipi* feature of percepts (he talks of ideas) to the superfluity of introspection: “Consciousness is not related to ideas as the stomach is related to the food that it takes in and digests. Indeed, it is ideas 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 (1925), p. 131]

misleading in suggesting an act/object structure of consciousness. And this structure we have been at pains to avoid; for talk of the act that is directed at its object is just talk of the inner eye, the inner sense, introspecting the contents of the mind. But it is, nevertheless, convenient to have a verb with which to describe the manner in which we enjoy our conscious mental states. I shall 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 should like to say, then, is that I feel the phenomenal redness of the tomato. And by saying that I feel the red quale I am not expressing any cognitive or doxastic relationship between me and the quale in question. This is the *thesis of the noncognitive nature of consciousness*.

The transition from the nonrelational account of sensation to the allegedly noncognitive nature of consciousness requires further comment. That consciousness is noncognitive is an important aspect of the present proposal; and it may be less than clear how this feature of consciousness drops out of the preceding account. My reason for wanting to insist on the noncognitive nature of consciousness is the fear that any construal of consciousness as cognitive will force me to readmit some kind of introspective mechanism into the account of consciousness. For in order for me to have any kind of epistemic relationship to my qualitative mental states I shall have to somehow represent them to myself, i.e., I shall have to introspect them in some sense of the term. This, then, is how one might motivate the transition from a nonrelational theory of experience to a noncognitive theory of consciousness. If one holds with Russell that “[t]here is a

duality which is essential in any form of knowledge...” and that “generally, knowing is distinct from that which is known” [Russell, 1975 (1959), p. 104] then the transition from nonrelational to noncognitive will not seem surprising. The account of sensation in terms of nonrelational percepts will have the upshot that percepts must be classified as noncognitive. 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 (1921), p. 141]

The subject or act was, of course, already exorcised on the joint grounds of unperceivability and economy. This lead 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 (1921), p. 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:

The closest conceivable relation between two objects is that of complete identity, so that in reality they are not two but one. 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; ...If such ideas were abandoned in the wake of scientific philosophy, it was because people became convinced that a union of the knowing consciousness with objects does not take place and indeed is not possible. 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 (1925), p. 81]

We can read this as a reaffirmation of Russell's claim that knowledge involves an essential duality. Thus, were this duality is missing knowledge is missing. This, Schlick holds, is the case in experience. Because we encounter this identity of subject object in the case of experience, experience must be counted as noncognitive.

[E]xperience 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; on cannot separate within it the sensing of the blue and the blue that is sensed. [Schlick, 1985 (1925), p. 197]

Because experience runs act and object into one we must hold that "[p]ure unelaborated perception or sensation is mere acquaintance⁸² (*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,

⁸² It is unfortunate that the translators chose the term “acquaintance” to render Schlick’s “*Kennen*” into English. For the term “acquaintance” immediately leads one to think of Russell’s *knowledge by acquaintance*. Before Russell accepted neutral monism he used the term “acquaintance” to stand for a certain kind of knowledge, i.e., for a cognitive relation the objects of which were, paradigmatically, sense data. After identifying the sense-datum with the sensing Russell discarded the notion of a sense-datum; consequently he had to give up the notion of knowledge by acquaintance. Schlick is, of course, at pains to deny that experience or sensation constitute knowledge of any kind.

1985 (1925), p. 87] Thus Russell and Schlick end up endorsing the thesis of 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 therefore consciousness is noncognitive.

REID'S CHALLENGE TO NONCOGNITIVISM

This argument for the noncognitive character of consciousness has been challenged in numerous ways. Using Schlick's mildly derogatory term, we can say that the "prophets of intuition" deny the first premise, viz., that knowledge is essentially relational. To follow up on this objection would take us too far afield. All those who uphold some version of the act/object analysis of experience challenge the second premise. One might want to defend the act/object analysis of experience for various reasons.⁸³ Instead of attempting to undermine them individually I have tried to present a rationale for embracing the neutral monist alternative to the act/object analysis of experience.

⁸³ The act/object analysis of experience had and has many able defenders. Its most celebrated defence may well be the one G.E. Moore gave it in his refutation of idealism. [Moore, 1903] It is worth noting that 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, p. 574] The same Russell later on felt sure, again on the grounds of purest empiricism, that the self or act 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. Chisholm 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, p. 317] Of course, this rationale for radical change in view cuts both ways; the seductive appeal of the new view is surely also motivated, at least in part, by its problem solving capacities.

But these two strategies do not exhaust all the possible objections to noncognitivism. I want to discuss a rival account that seems particularly troubling because it is so similar to the present noncognitivist account of consciousness in all but the crucial claim of the noncognitive nature of consciousness. Such a rival account is presented by Thomas Reid. The details of his theory are intricate and not always quite clear, but its main lines are easily discerned.

Reid distinguishes consciousness from introspection and insists that consciousness is independent of introspection. Introspection is rare and difficult to attain and not present in many. Consciousness of our mental operations, on the other hand, pervades our waking life. On the mental operation of sensation—the one that we have been concerned with throughout—Reid has this to say:

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 (no year), pp. 18-19]

And it is Reid who originally used the term “feeling” to talk about sensation so understood. Taken in this sense, the term “feeling” “has no object; the feeling and the thing felt are one and the same.” [Reid, 1785 (no year), p. 20] Up to this point Reid’s account and the one presented here are in perfect harmony. Surely we have established that Reid is not one to endorse an act/object analysis of sensation. But neither is Reid one of Schlick’s “prophets of intuition.” Unlike

those prophets he would reject any claim along the line that genuine knowledge requires one to become one with one's object. A brief consideration of his account of perceptual knowledge makes this clear. Reid distinguishes sensation from perception. Sensation, he tells us, "taken by itself, implies neither the conception nor belief of any external object." [Reid, 1785 (no year), p. 155] Perception, on the other hand, "implies an immediate conviction and belief of something external—something different both from the mind that perceives and from the act of perception." [Reid, 1785 (no year), p. 155] Through perception we come to know about the existence of external things. This act-object account of perceptual knowledge shows that Reid acknowledges the relational structure of (at least some forms of) knowledge. That is, we have reason to believe that Reid would accept Russell's claim that "generally, knowing is distinct from that which is known" [Russell, 1975 (1959), p. 104]. It is therefore incorrect to view Reid as just another "prophet of intuition."

So Reid seems to endorse both of the premises of the argument from which I—following Russell and Schlick—concluded the noncognitive character of consciousness. He endorses the relational character of knowledge; he endorses the nonrelational character of sensation or experience. Yet he holds that consciousness is cognitive. How does he make it work out?

As best I can determine, Reid gets off the train by denying that *all* knowledge is relational. Consciousness affords us what he calls an *immediate* certain knowledge of our mental operations, such as "our present pains, our pleasures, our hopes, our fears, our desires, our doubts, our thoughts of every kind." [Reid, 1785 (no year), p. 375] Though sensations are not explicitly mentioned in this list, Reid clearly thinks that they too are objects of the

immediate knowledge that consciousness affords. For in another place he tells us that “we are conscious of all our sensations.” [Reid, 1785 (no year), p. 186] Reid’s notion of the immediate knowledge of our sensations afforded us by our consciousness is best understood as a *nonrelational* form of knowledge. It would be wrong to think of the consciousness of our sensations to be something *in addition to* our feeling them. For Reid tells us that the consciousness that attends our sensations “is the evidence, the only evidence, which we have or can have of their existence.” [Reid, 1785 (no year), p. 24] If feeling the sensation were one thing and being conscious of it were another, then the consciousness of our sensations would *not* be the only way for our sensations to signal their existence to us. For in that case we could have evidence for the existence of our sensations either by feeling them or by being conscious of them. But in this case Reid’s claim that the consciousness of our sensations is the only way in which we can obtain evidence for their existence would turn out to be false. Feeling a sensation and being conscious of it must, therefore, be the same thing. And it is merely in virtue of our feeling or being conscious of our sensations that we come to have immediate certain knowledge of them. It would seem, then, that our knowledge of our sensations comes for free: for our sensations to exist is to be felt; for them to be felt is for us to be conscious of them; and for us to be conscious of them is for us to know them immediately and with certainty. No mental process over and above the sensation itself comes into play. There is no room in this story for reflection or introspection as the vehicle of immediate knowledge. Like our sensations themselves, our knowledge of our sensations is a nonrelational affair. On Reid’s account, to have a sensation is to know it immediately.

How is Reid's cognitive account of consciousness related to the noncognitivist account I have presented? How is this dispute to be resolved? At first sight it seems that the controversy must turn on one's resolving the substantial issue whether all knowledge is relational. But perhaps the stark opposition between Reid's cognitive and my noncognitive account of consciousness is more apparent than real. A remark that Chisholm made when confronted with a different but related difficulty points to a resolution of the dilemma. While discussing the directly evident, i.e., the bedrock of all epistemic justification, Chisholm raises the question of the epistemic status of the directly evident itself:

If we use our original characterization of the directly evident, we might think of the directly evident as that which "constitutes its own evidence," and therefore, in the terms of Sextus Empiricus, as that which is "apprehended through itself." But if we use the alternative characterization, we may think paradoxically of the directly evident as being that which is "evidence but not evident." In the one case, we are reminded of the prime mover that moves itself, and in the other, of the prime mover unmoved. [Chisholm, 1966, p. 30]

We have here one phenomenon under two significantly different descriptions. But no matter how you construe the evidential status of the directly evident, the directly evident is that which is evidence for all else. Similarly we might want to say that the cognitivist and the noncognitivist present two significantly different descriptions of the same phenomenon. But no matter how you construe the epistemic status of phenomenal consciousness, phenomenal consciousness is our inner light that shines without requiring an inner eye to perceive it.

It seems, then, that the dispute boils down to a choice: either you stick with the premise that all knowledge is relational and therefore involves some

form of inner representation. Then you have to go with the noncognitive theory of consciousness. Or you introduce a notion of nonrelational *immediate knowledge* and thus block the link between knowledge and introspection. The introduction of this special kind of knowledge that is peculiar to sensations (and maybe some other mental objects) allows you to hold a cognitive theory of consciousness without running the risk of having introspection sneaking into your account through the back door. I favor the noncognitive account because it strikes me as less mysterious. It does not take much to prophesy that Reid's notion of the immediate certain knowledge afforded by consciousness will be hard to tame. Given the arguments against the special knowledge thesis above, the claimed certainty must appear problematical. And in this case, where the knowing and the known are identical, it will be difficult to back off from the certainty claim gracefully. Where the knowing and the known are identical, there is no way to pry them apart sufficiently so as to allow the possibility of error to insert itself in the gap. Unlike the inner eye model of consciousness Reid's cognitive theory of consciousness does not hold out an easy route to fallibility. But even if it proved possible to domesticate Reid's theory and if, for some reason, we were forced to accept Reid's cognitive alternative, all would not be lost. We could still uphold the thesis of consciousness without introspection.

S U M M A R Y

In this Chapter I presented the neutral monist theory of consciousness as a way in which to understand the claim that to be conscious is to have qualia. Hopefully this claim is now no longer an empty slogan but, instead, an interesting and controversial thesis about consciousness.

An analysis of the root of the problems that plagued rival accounts of consciousness uncovered two assumptions a successful theory of consciousness has to do without. These were the act/object assumption and the subject/attribute assumption. I then discussed two theories that avoid these problematical assumptions in their respective ways.

The adverbialist theory of consciousness pulls the rug from under these assumptions by simply denying the existence of qualia as phenomenal properties. The question what bears the phenomenal properties does not so much as arise for the adverbialist. Thereby the subject/attribute assumption is dispensed with. And by reconceiving qualia as adverbial modifications of a subject's sensing states the adverbialist rids himself of the need for a qualia-bearing object. Thereby the act/object assumption is dispensed with.

Adverbialism was criticized on two counts. First, I argued that the functioning of the qualia turned adverbial modifications of sensing remains obscure. Second, I argued that the phenomenological adequacy of adverbialism is questionable.

I then proceeded to present and defend the neutral monist theory of consciousness. I argued that the neutral monist theory of experience conjoined with a double does of the bundle theory (the subjects as well as the objects of experience were reconstructed as bundles) provides a satisfactory account of consciousness. The identification of the act of sensing with the object sensed is the key to this account of consciousness. It yields the notion of the percept, i.e., the notion of a self-presenting object of experience. The bundle theory is the second important ingredient of the proposed account. First, it provides neutral monism with an account of what it is for a subject to have a quale. For a subject

to have a quale is for a percept to present itself to a subject, which it does by being a member of the group of percepts that the subject is. Second, it provides a satisfactory explanation of the relation between the objects of experience and its phenomenal properties. For a phenomenal property to inhere in an object of experience is for it to be a member of the bundle of phenomenal properties that the object of experience has. Thus we arrive at a fuller understanding of what it is to be conscious, of what it is to have qualia. The experiencer is understood as a bundle of percepts. The self-presenting percepts are conceived of as bundles of phenomenal qualities. Being conscious consists in there being bundles of the second kind among the element of bundles of the first kind. This is what I take to be the correct interpretation of the claim that to be conscious is to have qualia. This is also what I take to be the correct interpretation of the obscure claim the mistaken interpretation of which lies at the heart of introspectionism: *consciousness has built into it a directedness towards an apprehending mind.*

In the last section I defended the noncognitive nature of consciousness. Being conscious is not a matter of knowing, believing, or “seeing” something. This is a consequence of the fact that qualia present themselves to ourselves. To have qualia is to feel them. I ended by discussing Reid’s claim that we have immediate certain knowledge of our qualitative states and concluded that the dispute was primarily verbal.

A SOBERING LOOK AHEAD: A THEORY OF CONSCIOUSNESS IN SEARCH OF A METAPHYSICS

I wish I could stop at this point. But the neutral monist account of consciousness is seriously incomplete as it stands. I therefore have to at least point out the direction in which the completion of this project lies. To complete the account is more than I can do here and perhaps more than I can do anywhere. What I have presented so far is, so to speak, merely the form of a neutral monist theory of consciousness. What is missing is a comprehensive metaphysics rich enough not only to accommodate the neutral monist theory of consciousness but also the rest of what there is. Such a metaphysics exists. It is Neutral Monism.⁸⁴ Neutral Monism has not had many friends and the few there were have never fully articulated it. Thus the neutral monist theory of consciousness is burdened with a huge liability. Whether this liability is so great as to outweigh the many advantages of the theory is a judgement I shall have to leave to the reader.

I have tried to show that the neutral monist has an account of what it is for something to bear a quale, of what it is for a quale to inhere in what it qualifies. Put another way, the neutral monist has provided a home for qualia. The neutral monist also has an account of what it is for a subject to have a quale. The proposed account does not require one to introspect one's qualia, for the qualia present themselves to the subject who has them. And, finally, the neutral monist's account of what it is to have qualia is phenomenologically adequate. To

⁸⁴ The capitalization indicates that I am now speaking of the grand metaphysical theory; not of the fragment of the theory that I defended in this chapter.

have qualia in the neutral monist way is to be conscious in the sense that it is like something to have your qualia by having them present themselves to you. Neutral Monism passes Searle's trying on test.

Moreover I have tried to show that the neutral monist can give all these answers without getting caught up in the problems that plagued his rivals. Lets enumerate these problems and ask whether the neutral monist is susceptible to them.

Does the neutral monist have to attribute phenomenal properties to mental objects (like the dualist's sense data) whose existence is problematical/impossible? No.

Does the neutral monist have to attribute phenomenal properties to physical objects (like the relocationist's tomatoes) that cannot support them? No.

Does the neutral monist have to attribute phenomenal properties to mere possibilia that either cannot support them (like, for example, the retreating relocationist's abstract states of affairs) or if they can, are "so distant" from us (like the retreating relocationist's Lewisian counterparts) that we cannot have the qualia they support? No.

Does the neutral monist have to transmogrify qualia into something unrecognizable (like the adverbialist's adverbial modifications of sensings) the having of which is quite likely to be like nothing? No.

Does the neutral monist have to rely on mental acts (like, for example, the introspectionist's introspection) the exercise of which makes the having of qualia be like nothing? No.

Does the neutral monist have to compromise the phenomenologically given features of consciousness (like, for example, the eliminativist who seems to feign general anesthesia) when trying on his theory? No.

In cataloging the numerous advantages of the neutral monist theory of consciousness I have tried to present this theory in the best possible light. A clear awareness of neutral monism's many benefits will be necessary if those benefits are not to be outbalanced by the theories murkier aspects. On the negative side I see one big problem: the metaphysics of it all are obscure.

My remarks on the metaphysics of neutral monist theory of consciousness will be short and, I fear, somewhat superficial. I choose this route not because the questions are not deserving of a fuller discussion. On the contrary, the questions are fascinating. But an adequate treatment of these issues would require a book of its own. I believe that the neutral monist theory of consciousness must be complemented by a metaphysics that places the apparatus on which this theory relies—qualia, self-presentation, the bundle theory of persons, the bundle theory of sensible objects—into a broader context of a general metaphysical theory of the world. As I shall indicate below, I believe that materialism does not have the resources to accommodate all this machinery. Dualism is much less restrictive in what it allows to exist. But the facileness with which dualism accommodates the ontological baggage of the neutral monist theory of consciousness is also what robs dualism of its explanatory power. The most promising background metaphysics for the Neutral Monist theory of consciousness is a full blown neutral monist metaphysics along the lines developed by Russell.

Lets begin by asking: How nonmaterialist is the neutral monist theory of consciousness? At the beginning of this Chapter I said that we want unrelocated qualia. That is, we want qualia that qualify mental particulars. If the mind is the brain, this leaves open the possibility that qualia are properties of the brain. But as the discussion of the relocation thesis has shown, I take a dim view of the possibility of reducing qualia to respectable physical properties. Thus it would seem that qualia come at the price of property dualism. On the other hand it may seem that this is all the dualism we will get. For the neutral monist approach was marked by a distaste for substances. Thus Cartesian substance dualism does not seem to be in the offing.

I fear, however, that this simple attempt to banish the danger of substance dualism is hollow. It is true that the neutral monist theory of consciousness does not say “there are nonphysical substances.” But this is simply so because neutral monism does not talk about substances at all. And in its own way it postulates the existence of particulars the nature of which should be no less offensive to the materialist than are Cartesian substances. For the neutral monist’s percepts are particulars that are wholly constituted of phenomenal, i.e., nonphysical, properties. If physical properties—like, for example, having an atomic structure—were allowed into a percept, then percepts would not agree with phenomenal properties any better than do material substances. I am aware that pointing out *one* physical property that percepts cannot have is not to show that percepts cannot have any physical properties. Percepts do have some properties that, traditionally, have been considered marks of the nonmental. For example, percepts are extended. But should this be any consolation to the materialist who considers the neutral monist theory of perception? Does the percept’s being

extended make them into respectable denizens of the physical brain? I think not. One does not get a materialistically respectable particular by investing it with some arbitrary selection of physical properties. An extended particular that aspires to be material will have to have all the properties that go with this station. The way we currently think about physical particulars, this will include such properties as having an atomic or quantized structure. And these properties cannot go into a bundle that is a percept.⁸⁵

Another consideration persuades me that materialism cannot be the right background metaphysics for the neutral monist theory of consciousness. If percepts are extended qualified particulars in my brain then there should be phenomenally red particulars in my brain upon my experiencing the redness of a ripe tomato. But this is not so. This is a very plain consideration, but I cannot see what is wrong with it.⁸⁶

If materialism cannot hold the correct theory of consciousness materialism must go. That is, I am willing to sacrifice materialism on the altar of phenomenological adequacy. But is dualism the obvious choice for an alternative metaphysics? The nonphysical realm of the dualist provides what seems like a bottomless receptacle for everything that seems problematical in the neutral monist theory of consciousness (problematical, that is, from a materialist point of

⁸⁵ It seems that Sellars's *sensa* of the completed scientific image are just such alarmingly emaciated "scientific" objects: extended but otherwise unstructured (i.e., grainless) physical objects. Perhaps some future physics will be accommodating enough to allow for the existence of percepts understood along Sellarsian lines. But until this future time is here, percepts don't seem to fit into the physical image of the day.

⁸⁶ In chapter 4 I used the term "colored brain hypothesis" to refer to a view like the one I just rejected. There I argued that there are ways of defending something akin to the colored brain hypothesis. The lines of defence adumbrated there are not available to the materialist.

view). Irreducibly mental properties are, of course, no problem. The colored brain problem does not arise; for percepts are not in the brain but in the mind. And the extended percepts are located in a mental space (that is distinct from the physical space) that also contains the subject. I am reluctant to accept dualism as a metaphysical framework for the neutral monist theory of experience. This reluctance does not stem from any special difficulties that arise from combining dualism with the neutral monist theory of experience. As far as I can tell there are no such special problems. It is my general disenchantment with Dualism as an explanatory theory that leads me to reject it.

The metaphysics that offers the most promise for delivering a unified and comprehensible account of the place of consciousness and its place in nature is Neutral Monism. Neutral Monism is a reductive metaphysics because it holds that everything is ultimately composed of particulars of the kind with which we become acquainted in conscious experience, i.e., percepts. It is a monism inasmuch as it holds that the particulars “that make a living brain are actually identical with those that make the corresponding mind.” [Russell, 1956, p. 147] And it is neutral inasmuch as the particulars are themselves neither mental nor physical. Mental and physical events are (logically) late arrivals on the scene. Both are composed of the same kinds of particulars, differently grouped. Typical Leibniz Law objections are avoided by elaborating the principles of grouping that yield physical and mental events respectively. Concerning the manner in which percepts go into the making of the mind/brain Russell has this to say:

I do not think that my visual percepts are a “portion” of my brain; “portion” is a material concept. ...Observe that a “portion” of a brain is a set of points (or minimum volumes); an event may be a member of certain points (or minimum volumes) that are members of the brain, and is then said to be “in” the brain, but it is not “part” of the brain. It is a member of a member of the brain. [Russell, 1946, p. 705-706]

This is to be taken quite literally. The objects (like, for example electrons) that compose your brain in the physical order of things are themselves composed (at least in part) by your percepts. As for matter that is not part of a living brain, it too is composed of similar particulars. We happen to know nothing about the intrinsic nature of these particulars. But there is no reason for believing that these particulars are intrinsically totally unlike the ones that are our percepts.

The Neutral Monist picture allows for unrelocated, unreduced, unborne, self-presenting qualia and percepts. And it does not accommodate all this “weird stuff” by shunting it (dualism style) into some peculiar nonphysical realm, the existence of which causes more problems than it solves. Nor does it try to accommodate qualia (materialism style) by initially going easy on the phenomenology and by subsequently debunking qualia as simple being ordinary physical items not apprehended as such. In the Neutral Monist inventory of the world qualia and percepts feature as basic, unreduced neutral (*not mental!*) particulars that are equidistant from the constructs we call mind and matter.

This is the barest sketch of Neutral Monism. It accommodates percepts by interpreting them as the basic elements from which all other categories are to be constructed. By making percepts basic in the order of being it guarantees conscious experience a secure place in Nature. Perhaps it is possible to integrate the neutral monist theory of consciousness into a less adventurous metaphysics.

But since I cannot see how to do this, I have to admit that the neutral monist theory of consciousness as I proposed it is saddled with an enormous IOU.

REFERENCES

- 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 and Norman Malcolm. 1984. Consciousness & Causality. A Debate on the Nature of Mind. Oxford: Basil Blackwell.
- 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. 1980. The Nature of Mind. Brighton, Sussex: Harvester Press.
- Armstrong, D.M. 1961. Perception and the Physical World. London: Routledge & Kegan Paul.
- Aune, Bruce. 1967. Knowledge, Mind, and Nature. Atascadero: Ridgeview Publishing Company.

- Ayer, A.J. 1971. Russell and Moore. The Analytical Heritage. London: Macmillan.
- Berkeley, George. 1985 (1713). Three Dialogues Between Hylas and Philonous. 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. 1978. "Troubles With Functionalism." Originally in [Savage, 1978]. Reprinted in [Block, 1980], Vol. 1. Pages: 268-305. Page references are to this reprinting.
- 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.
- Brockman, John, ed. 1990. Speculations. New York, NY: Prentice Hall 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. Second Edition. Notre Dame: University of Notre Dame Press.
- Carruthers, Peter. 1989. "Brute Experience." The Journal of Philosophy 86. Pages: 258-269.
- Castañeda, Hector-Neri. 1979. "Philosophical Method and Direct Awareness of the Self." In [Sosa, 1979]. Pages: 1-58.
- Chisholm, R.M. 1973. "Homeless Objects." In [Chisholm, 1982]. Pages: 37-52.
- Chisholm, R.M. 1979. "Objects and Persons: Revisions and Replies." In [Sosa, 1979]. Pages: 317-388.
- Chisholm, R.M. 1983. "Believing as an Intentional Concept." In [Parret, 1983]. Pages: 48-56.

- 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. 1982. "Beyond Being and Nonbeing." In [Chisholm, 1982]. Pages: 53-67.
- Chisholm, Roderick M. 1982. Brentano and Meinong Studies. Atlantic Highlands: Humanities Press.
- Churchland, Patricia S. 1983. "Consciousness: The Transmutation of a Concept." Pacific Philosophical Quarterly 64. Pages: 80-95.
- Churchland, Paul. 1979. Scientific Realism and the Plasticity of Mind. Cambridge: Cambridge University Press.
- Colodny, Robert, ed. 1962. Frontiers of Science and Philosophy. Pittsburgh: Pittsburgh University Press.
- Cummins, Robert. 1984. "The Mind of the Matter." PSA 2. Pages: 791-798.
- Davidson, R.J., G.E. Schwarz and D. Shapiro, eds. 1983. Consciousness and Self-Regulation, Volume 3. Plenum Publishing Corporation.
- Davidson, R.J., G.E. Schwarz and D. Shapiro, eds. 1983. Consciousness and Self-Regulation, Volume 3. Plenum Publishing Corporation.
- 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. 1982. "How to Study Human Consciousness Empirically or Nothing Comes to Mind." Synthese 53. Pages: 159-180.
- Dennett, Daniel C. 1987. "Consciousness." In [Gregory, 1987]. Pages: 160-164.

- 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., ed. 1981. Brainstorms. Philosophical Essays on Mind and Psychology. Brighton, Sussex: Harvester Press.
- Dennett, Daniel. 1987. The Intentional Stance. Cambridge, Mass.: MIT Press.
- Dennett, Daniel. 1987a. The Intentional Stance. Cambridge, Mass.: MIT Press.
- Dennett, Daniel. 1987b. "Setting off on the Right Foot." In [Dennett, 1987]. Pages: 1-11.
- Dennett, Daniel. 1991. Consciousness Explained. Boston: Little, Brown and Company.
- Dretske, Fred I. 1969. Seeing and Knowing. Chicago: The University of Chicago Press.
- Eddington, Arthur Stanley. 1929. The Nature of the Physical World. New York: Macmillan.
- Farrell, B.A. 1950. "Experience." Mind 59. Pages: 170-198.
- Feigl, Herbert. 1967. "Postscript After Ten Years." In [Feigl, 1967]. Pages: 135-169.
- Feigl, Herbert. 1967. The Mental and the Physical. The Essay and the Postscript. Minneapolis: University of Minnesota Press.
- Flanagan, Owen. 1991. The Science of Mind. Second Edition. Cambridge, Massachusetts: 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. God and the Soul. London: Routledge and Kegan Paul.
- Geach, Peter. 1969. "Immortality." In [Geach, 1969]. Pages: 17-29.
- 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. 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.
- 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, ed. 1981. Mind Design. Philosophy, Psychology, Artificial Intelligence. Cambridge, MA: MIT Press. A Bradford Book.
- Haugeland, John. 1981. "Semantic Engines." In [Haugeland, 1981]. Pages: 1-34.
- Hess, Peter H. 1988. Thought and Experience. Toronto: University of Toronto Press.
- 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.
- Hume, David. 1973 (1739). A Treatise of Human Nature. 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.
- 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.
- Kneale, William. 1969. "Review of Armstrong's *A Materialist Theory of Mind*." Mind 78. Pages: 292-301.
- Lehrer, Keith. 1990. Theory of Knowledge. Boulder and San Francisco: Westview Press.

- Lloyd, Dan. 1989. Simple Minds. Cambridge, MA: MIT Press. A Bradford Book.
- Locke, John. 1982 (1689). An Essay Concerning Human Understanding. Oxford: Oxford University Press.
- Lockwood, Michael. 1989. Mind, Brain and the Quantum. The Compound 'I'. Oxford: Basil Blackwell.
- 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.
- Lyons, William. 1986. The Disappearance of Introspection. Cambridge, MA: MIT Press. A Bradford Book.
- Marcel, A. and E. Bisiach, eds. 1988. Consciousness in Contemporary Science. Oxford: Oxford University Press.
- 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.
- McGinn, Colin. 1982. The Character of Mind. Oxford: Oxford University Press.
- McGinn, Colin. 1989. "Can We Solve the Mind-Body Problem?" Mind 98. Pages: 349-366.
- McGinn, Colin. 1991. 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.
- Moore, G.E. 1903. "The Refutation of Idealism." Mind 12. Pages: 433-453. Reprinted in [Moore, 1951 (1921)]. Pages: 1-30. Page references are to this reprinting.

- Moore, G.E. 1951 (1921). Philosophical Studies. New York: The Humanities Press.
- Moore, G.E. 1951 (1921). Philosophical Studies. New York: The Humanities Press.
- Moore, G.E. 1966. Some Main Problems of Philosophy. 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.
- Natsoulas, Thomas. 1982. "Conscious Perception and the Paradox of "Blind Sight"." In [Underwood, 1982], Volume 3. Pages: 79-109.
- Parret, H., ed. 1983. On Believing. Berlin: DeGruyter.
- Parsons, Terence. 1987. "Entities without Identity." Philosophical Perspectives 1. Pages: 1-19.
- 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.

- Pollock, John L. 1989. How to Build a Person. A Prolegomenon. Cambridge, MA: MIT Press. A Bradford Book.
- Pollock, John. 1986. Contemporary Theories of Knowledge. Totowa: Rowman & Littlefield.
- Posner, Michael I., ed. 1989. Foundations of Cognitive Science. Cambridge, Mass: MIT Press.
- Posner, Michael I., ed. 1989. Foundations of Cognitive Science. Cambridge, Mass: MIT Press.
- Price, H.H. 1954 (1932). Perception. London: Methuen & Co.
- Radner, Daisie and Michael Radner. 1989. Animal Consciousness. Buffalo: Prometheus Books.
- Reid, Thomas. 1785 (no year). Essays on the Intellectual Powers of Man. Edited and abridged by A.D. Woozley. Charlottesville: Ibis Publishing.
- Rey, Georges. 1983. "A Reason for Doubting the Existence of Consciousness." In [Davidson, et al., 1983], Volume 3. Pages: 1-39.
- Russell, Bertrand. 1912. "Review of William James *Essays in Radical Empiricism*." Mind 21. Pages: 571-575.
- 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 (1956)]. 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 (1927). The Analysis of Matter. New York: Dover.
- Russell, Bertrand. 1956. "Mind and Matter." In [Russell, 1956]. Pages: 135-153.
- Russell, Bertrand. 1956. Portraits from Memory. London: George Allen & Unwin.

- Russell, Bertrand. 1975 (1959). My Philosophical Development. London: George Allen & Unwin.
- Russell, Bertrand. 1977 (1956). Logic and Knowledge. Essays 1901-1955. Edited by R.C. Marsh. London: George Allen & Unwin.
- Russell, Bertrand. 1978 (1921). The Analysis of Mind. London: George Allen & Unwin.
- Russell, Bertrand. 1980 (1940). An Inquiry into Meaning and Truth. London: George Allen & Unwin.
- 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. 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 (1925). General Theory of Knowledge. LaSalle: Open Court.
- Searle, John R. 1980a. "Author's Response." The Behavioral and Brain Sciences 3. Pages: 450-456.
- Searle, John R. 1980b. "Minds, Brains, and Programs." The Behavioral and Brain Sciences 3. Pages: 417-424.
- 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.

- 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.
- Stoppard, Tom. 1972. Jumpers. New York: Grove Press.
- Tomberlin, James E., ed. 1990. Action Theory and the Philosophy of Mind. Philosophical Perspectives, Volume 4. Atascadero: Ridgeview.
- Underwood, Geoffrey, ed. 1982. Aspects of Consciousness, Volume 3. London:
- 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.
- Wisdom, John. 1963 (1934). Problems of Mind and Matter. 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.