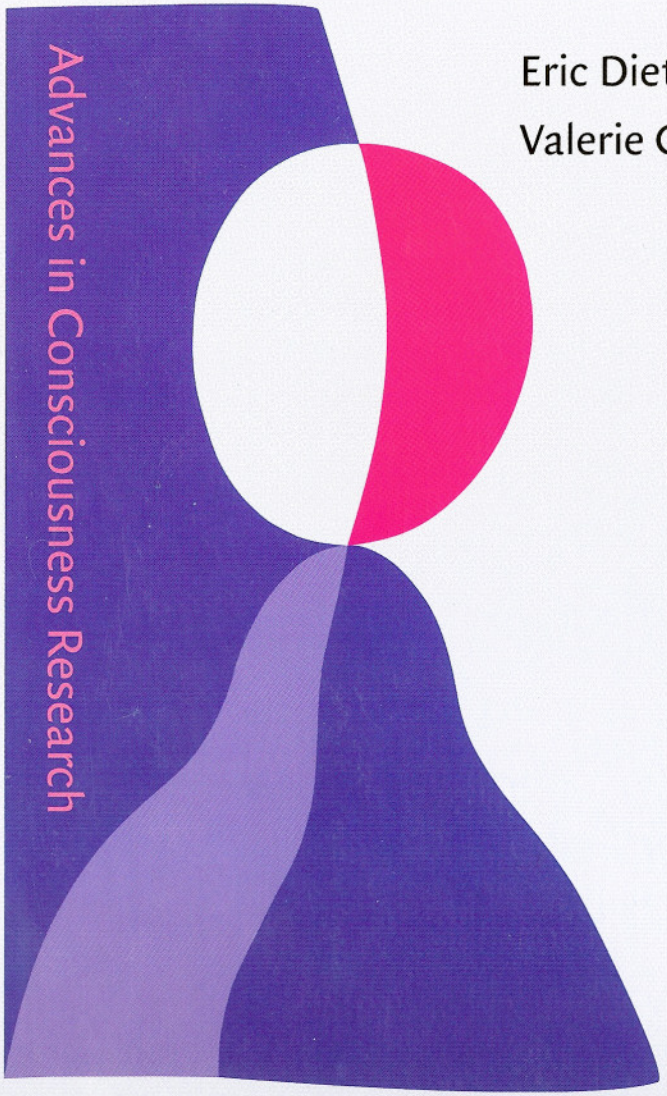
Sisyphus's Boulder

Consciousness and the limits of the knowable



Eric Dietrich and
Valerie Gray Hardcastle

Consciousness lies at the core of being human. Therefore, to understand ourselves, we need a theory of consciousness. In *Sisyphus's Boulder*, Eric Dietrich and Valerie Hardcastle argue that we will never get such a theory because consciousness has an essential property that prevents it from ever being explained. Consequently, philosophical debates over materialism and dualism are a waste of time. Scientific explanations of consciousness fare no better. Scientists do study consciousness, and such investigations will continue to grow and advance. However, none of them will ever reveal what consciousness is. In addition, given the centrality of consciousness in philosophy, Dietrich and Hardcastle claim that philosophy itself needs to change. That the central problems of philosophy persist is actually a profound epistemic fact about humans. Philosophy, then, is a limit to what humans can understand.

ISBN 90 272 5196 7 (Eur) / 1 58811 602 6 (US)

John Benjamins Publishing Company

Advances in Consciousness Research

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

Series A: Theory and Method. Contributions to the development of theory and method in the study of consciousness.

Volume 60

Sisyphus's Boulder: Consciousness and the limits of the knowable
by Eric Dietrich and Valerie Gray Hardcastle

Sisyphus's Boulder

Consciousness and the limits of the knowable

Eric Dietrich
Binghamton University

Valerie Gray Hardcastle
Virginia Polytechnic Institute and State University

John Benjamins Publishing Company
Amsterdam/Philadelphia

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

Library of Congress Cataloging-in-Publication Data

Eric Dietrich
Sisyphus's Boulder : Consciousness and the limits of the knowable / Eric Dietrich and Valerie Gray Hardcastle.
p. cm. (Advances in Consciousness Research, ISSN 1381–589X ; v. 60)
Includes bibliographical references and indexes.
1. Consciousness. I. Hardcastle, Valerie Gray. II. Title. III. Series.

B808.9.D54 2004
126--dc22 2004059570
ISBN 90 272 5196 7 (Eur.) / 1 58811 602 6 (US) (Hb; alk. paper)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

To my wife: Tara
ESD

To my children: Kiah, Cheshire, and Quinn
VGH

Table of contents

Acknowledgements

Several people have helped in this endeavor. Some directly; some indirectly through their works. From the second set, we have been most heavily influenced by Kathleen Akins, Ned Block, David Chalmers, Patricia and Paul Churchland, René Descartes, Owen Flanagan, Jerry Fodor, George Graham, Terry Horgan, Joe Levine, Thomas Nagel, Colin McGinn, Phil Merikle, Graham Priest, John Searle, Sydney Shoemaker, William Wimsatt, the Association for the Scientific Study of Consciousness, and the McDonnell Neurophilosophy Fellows.

From the first set, we thank the participants of a terrific, May 2001 workshop on machine consciousness, held at Cold Spring Harbor Labs; their comments on an early draft of some of the central ideas in this book were quite useful. We thank David Chalmers for discussions about some of the arguments we use. We thank our students in various classes on whom we have tried out these ideas over the last several years, especially Hardcastle's 1999 graduate seminar on the scientific study of consciousness. Thanks also go to Jerry DeJohn, Thony Gillies, and Güven Güzeldere for working with us on some of the ideas central to this book. We thank Brian McKeon, Michael Winkley, and Sara Doane for proofreading the final draft (it only seems fair that any remaining errors should be blamed on them), and Maxim Stamenov, the editor of the Advances in Consciousnes Research series, and Bertie Kaal at John Benjamins Press, for their unending patience as we wrote and rewrote this book more times than is healthy.

Many of the ideas found in this book had their origins in previous publications. The most directly relevant are: Dietrich, E. (1998), "Zombies only seem logically possible: or How consciousness hides the truth of Materialism: A Critical Review of *The Consciousness Mind* by David Chalmers," *Minds and Machines, 8* (3), pp. 441–461; Dietrich, E. (1999), "Fodor's gloom, or What does it mean that dualism seems true?" *Journal of Experimental and Theoretical Artificial Intelligence, 11* (2), 145–152; Dietrich, E. & A. Gillies (2001), "Consciousness and the limits of our imaginations," *Synthese, 126* (3), pp. 361–381 (permission to reprint sections of this paper in the appendix to this

work has been graciously granted by Kluwer Academic Publishers, Synthese, and my wonderful colleague, Thony Gillies); Güzeldere, G., Flanagan, O., & Hardcastle, V. G. (1999). "The Nature and Function of Consciousness: The Case of Blindsight." In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences*, Second Edition (pp. 1277–128). Cambridge, MA: The MIT Press; Hardcastle, V. G. (1996), "The Why of Consciousness: A Non-Issue for Materialists," *Journal of Consciousness Studies, 3* (1), pp. 7–13; Hardcastle, V. G. (2000). "How to Understand the N in NCC." In T. Metzinger (Ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (pp. 259–264). Cambridge, MA: MIT Press.

Finally, and most especially, we thank Neo, Trinity, Morpheus, Agent Smith, and the whole gang in *The Matrix* for making the Cartesian Intuition so compelling. Descartes would have been proud.

Introduction

The problem of consciousness is completely intractable. We will never understand consciousness in the deeply satisfying way we've come to expect from our sciences. This is due to a logical property of consciousness itself. The intractability of consciousness defines a boundary to knowledge; it limits what we can know about consciousness. Because it is a barrier, it does not entail dualism or idealism, nor is it a threat to materialism. At the same time, this property does explain both the resurgent plausibility of dualism as well as the fact that very little scientific progress to date has been made on reductively explaining consciousness.

We came to these conclusions when we sat down one afternoon several years ago to discuss the problem of intractability and the argument from it to consciousness's explanatory dead-end. Dietrich argued from the dead-end to the conclusion that there would not be a science of consciousness worthy of the name. Hardcastle argued from the fact that there already is a science of consciousness and the near-certainty that it will grow in robustness to the conclusion that there will somehow be an explanation of consciousness, of some sort. Neither one of us appreciated the value of the others' viewpoint, to say the least. Yet both of us remained convinced of the correctness of the intractability, so we continued talking.

A key moment came when we realized that sciences don't have to explain things – at least they don't have to explain things in a way that connects them with that vast realm of explanations comprising science's great successes. Once we realized this, we realized that the focus of debate in consciousness research was not on target. The (or a) primary focus is between those who think that consciousness is a nonmaterial property of the universe and those who think that it is a material property – roughly, the debates about consciousness revolve around dualism versus materialism. But this is not the right focus, given consciousness's intractability. The right focus is between those who believe that consciousness will one day be explained satisfactorily and those who deny that it will. The real debate, in other words, is over whether consciousness will occupy a unique place in the constellation of science, or whether standard science

will one day grow to enfold consciousness and make it understandable. In this book, we argue that consciousness occupies a unique place within science, and that, in a deep sense, consciousness will always remain beyond understanding. And this fact is due to an under-appreciated property of consciousness itself.

Yet understanding is somewhat like comfort: it is bred by familiarity. Since there will likely be a robust but nonexplanatory science of consciousness that will traffic in manipulations of various sorts, and since use begets familiarity, a kind of understanding of consciousness will grow – its deeper, more mysterious aspects will come to seem corralled . . . and that will be a kind of understanding. Which is good, because this will have to do. (We point out, too, that understanding why consciousness is in certain ways beyond understanding is itself a kind of understanding. This will add to our increasing sense that science is, over time, reigning in consciousness, and that we understand it.)

After working out all of this, we then realized that the nearly complete intractability of the problem of consciousness changes the way philosophers should understand philosophy. This may seem too big of a leap. Understanding consciousness is in part a philosophical problem because it involves basic, purely conceptual matters; there's a logic involved when wrestling with consciousness. But there is something else – an insight due to Thomas Nagel. The problem of consciousness involves attempting to reconcile two conflicting points of view: the subjective and the objective. And this attempted reconciliation is, according to Nagel, shared in common by *all* philosophy problems. The intractability property of consciousness says, basically, that the subjective point of view is logically irreducible to the objective. Assuming that going the other way – reducing the objective to the subjective – is also not in the cards, then the two are logically irreconcilable. If all philosophy problems are at bottom attempts to reconcile these two points of view, and they are irreconcilable in principle, then philosophy is a Sisyphean task. Or at the very least, philosophy has a Sisyphean component.

Philosophy, then, on our view, is not a set of deeply posed problems that we are solving, albeit slowly (glacially so). Instead, the persistence of the problems of philosophy emerges as a profound *epistemic* fact about humans. But not just humans. We argue that any conscious cognitive agent or being is going to be beset by these problems, and beset by them forever.

So, in short, we claim that the profound intractability of the problem of consciousness shows that fundamental philosophical mysteries of the universe are epistemological, rather than metaphysical. And this changes everything about how we should understand philosophy.

We argue for all this in this book. In Chapter 1, we frame the debate, altering it from the way it is usually laid out. We couch the discussion as a debate between *naturalists* and *mysterians*. Naturalists believe that there will one day be a robust, explanatory theory of consciousness; mysterians claim that science is virtually powerless in the face of consciousness. They are both wrong. In Part I, we argue that naturalism is wrong. In Part II, we argue that mysterianism is wrong. In the final part, we show how our approach to the problem of consciousness changes the debate in philosophy from metaphysics to epistemology.

Our goal in this book is to end fruitless debate about the metaphysical status of consciousness, and to clear the way for a decent science of consciousness by removing from it any guilt it may have at not explaining consciousness. Nothing can explain consciousness. So it can't be a failing of science that it can't explain it. The epistemic fact we have to come to terms with is that what lies at the core of our being, the core of our agenthood, is something that lies, not beyond the limit of our science, but beyond the limit of knowledge.

Chapter 1

Intuitions at an impasse

The explanatory landscape

> At the very end of his long effort measured by skyless space and time without depth, the purpose is achieved. Then Sisyphus watches the stone rush down in a few moments toward that lower world whence he will have to push it up again toward the summit. He goes back down to the plain.
>
> *Albert Camus*

1.1 Consequences of some obvious but under-appreciated facts

1.1.1 What is consciousness?

Consciousness is the way the world seems to us, the way we experience it, feel it. Taste an onion, see a rainbow, smell a dead skunk on a hot summer's day, stub your toe on the foot of the bed frame at 4:00 A.M., hear your dog breathe or a baby gurgle and coo. These are experiences, bits of our phenomenology, and it is these experiences that somehow give us our subjective point of view. (We will sometimes use the term "qualia" to refer to these qualitative feels.) We have experiences because we are conscious. Or, rather, our having them constitutes our being conscious. Being conscious is what makes it fun or horrible or merely boring to be a human. Using the phrase that Thomas Nagel (1974) made famous, we can say that a being is conscious if there is something it is like to be that being.

The above ostensive definition of consciousness, appealing to something we can only assume our readers have, will have to do, for there is no more robust scientific, third-person definition available. And this ostensive definition is the same one researchers have been using to define consciousness since research of any sort was first conducted on it. This fact is well known and has been extensively commented upon in the literature. But we think it indicates something not appreciated or at least not generally acknowledged: Research

on consciousness is not getting any closer to a satisfying explanation of consciousness.

Conscious experience is the most familiar thing in the world – to you. Descartes was right: you know nothing as certainly as your own conscious experiences and, next to that, nothing as certainly as the fact that you are conscious.[1] This is quite odd, given that the inexorable march of science has somehow managed to say little illuminating about consciousness. This may be one of the many spots where the reader will disagree with us, but it is true nonetheless. If *science* – and not just a few individual scientists here and there – had managed to say something substantial about consciousness, then there would now be some agreement on a nonostensive, more theoretically-based definition of consciousness, as well as how to reductively explain it. Compare the evolution of life or the behavior of masses in a gravitational field. Science has said something illuminating about these phenomena, and that is why there is agreement about them. But there is absolutely no agreement about consciousness, neither concerning what it is nor how to explain it.

So, oddly, the thing we as individuals know best – our conscious experience – is the thing about which we as a collective of understanders know least. As Bertrand Russell remarked, "The sciences have developed in an order the reverse of what might have been expected. What was most remote from ourselves was first brought under the domain of law, and then, gradually, what was nearer: first the heavens, next the earth, then animal and vegetable life, then the human body, and last of all (as of yet imperfectly) the human mind" (1961). Human consciousness falls at the very end of our list of intellectual challenges. There is a good chance that it will remain on that list, forever a challenge, forever unexplained.

This is an important point. Our whole enterprise begins with it, with this realization: Consciousness is a big, deep mystery. It is completely surprising, both that it exists and that it exists the way it does. Why hasn't science had anything useful to say about consciousness? Our modern scientific sentiments tell us that a complete and detailed catalog of the physical, chemical, biological, and computational facts of the world should entail both the existence of consciousness and its particular properties. We don't have such a catalog yet. But even so, one would have expected that our current admittedly incomplete but still quite good catalog would have at least gestured toward consciousness, much as it gestures towards quantum computers or a multiverse. But it does not. In fact, our current scientific knowledge does not in any way entail, suggest, or even hint at the existence of consciousness. Yet there it is in all its glory. Consciousness is utterly strange. [2]

The explanatory problem of consciousness is the problem of providing a good third-person, i.e., *objective*, account of consciousness – an account that reduces consciousness to some brain functioning, perhaps, or to some higher-level cognitive or perceptual functioning. Every human (presumably) understands at least something of his or her own consciousness from a first-person perspective, but what is needed to solve the explanatory problem of consciousness is a third-person perspective. The two fundamental questions such an explanation would allow researchers to answer are (1) why does consciousness exist – why is it like something to be a human, or any other conscious being, and (2) why do we, or any conscious being, have the particular experiences we have – why does biting into an onion lead to the particular experiences that it does and not to experiencing, say, blue?

These are very difficult questions, and no one has anything remotely approaching a convincing answer to them. (Of course, many researchers think they have the correct answers to questions (1) and (2), but they can't convince very many others that they have the correct answers. If others don't find an argument compelling, perhaps it isn't.) We flatly admit that we don't know how to answer (1) and (2). However, we have the next best thing: We have an explanation for why answering them has proven elusive, and our explanation has a surprising and large effect on any future science of consciousness. Not knowing how to answer questions (1) and (2) is *not* a bar to having a science of consciousness. Science, properly understood, divorces answering these questions from theorizing about the phenomenon.

Part of the key to appreciating this point is recognizing that scientific theories need not be explanatory in a full, satisfying way that comports with our folk views of, or our more sophisticated views of, or even our deeply held intuitions of, how the world works. In many sciences, having a theory means having an explanation (e.g., germs in medicine, plate tectonics in geology, natural selection and genetics in biology). But this isn't true in all sciences. And it is especially *not* true in the study of consciousness. Our position, then, is that a science of consciousness is in the offing, but crucial questions about consciousness will remain unanswered because they are necessarily unanswerable, and hence, an explanation of consciousness will not be forthcoming. We therefore distinguish between a science of consciousness and an explanation of consciousness.

1.1.2 But what is consciousness really? An embarrassment of explanations

Many people are quite sanguine about the prospects of fully explaining consciousness. In fact, many find consciousness not very mysterious at all. Consciousness really is just *X*, they say, and *X* can be explained scientifically. Therefore, a scientific explanation of consciousness is just around the corner. Others say, consciousness really is just *X* and we can't explain *X* scientifically, but we can explain it some other way, religiously, for example. Therefore, while a scientific explanation of consciousness is impossible, we will nevertheless have another, different kind of explanation.

We will concentrate just on the scientific attempts at explaining consciousness. Scientists have hypothesized many intriguing and insightful things as being identical to consciousness. Here are some of them: attention, autobiographical memory, being awake, body-based perspectivalness, neural competition, episodic memory, executive processing, feedback, feature integration, 40 Hz neural oscillations in human brains, high-level encoding, intentionality (as in intending to do something), intentionality (as in a mental representation's being about something in the world), meta-processing, mind-based perspectivalness, quantum effects in the microtubules of neurons, recursivity, reflective self-awareness, reportability, salience, sense of self.

That is quite a list. Note how far and wide it ranges. Some of these suggestions reduce consciousness to brain processes (40 Hz oscillations, quantum effects in microtubules), others attempt to identify consciousness with some psychological process or property (attention, executive processing, various memory systems), while still others move the problem of consciousness to some other cohort property of equal rank (intentionality, reflective self-awareness, subjectivity, sense of self). None have succeeded. All the items on the list have one of two properties: they either are necessary (at best), but not sufficient, for consciousness, or are as puzzling as consciousness itself. All of the attempts to reduce consciousness to brain processes or to psychological processes give us necessary properties for consciousness (perhaps). All the attempts to explain consciousness by using some mental property of equal rank merely replace the problem of consciousness with one equally puzzling.

For example, the suggestion that consciousness is captured by populations of neurons oscillating at 40 Hz is an attempt to identify consciousness with a better understood brain process. If person P is conscious, then P's brain exhibits the appropriate 40 Hz oscillations. But other mental processes might also be associated with this 40 Hz oscillation, and indeed probably are. We are missing something crucial: we have no way of articulating why it is that

40 Hz oscillations and not something else co-occurring with these oscillations is identical with consciousness. What we want is what we have in, say, thermodynamics. In thermodynamics, temperature in solids, liquids, and gases (as measured by some device, not how hot or cold something *feels*) is identified with the motion of molecules – temperature is *reduced* to molecular motion. But the reduction is done *without remainder*, which is to say that molecular motion is both necessary and sufficient for temperature, or that temperature logically supervenes on molecular motion. There is nothing to temperature other than molecular motion: given a certain amount of molecular motion, a certain temperature results. The same cleanliness of connection eludes us in the oscillation case (and in the microtubule case, and in any other attempt to reduce consciousness to some other process).

Reducing consciousness to psychological properties fares no better. For example, reducing consciousness to executive processing and control leaves us in the dark about what it is about executive control that entails consciousness. The same is true for intentionality. To the extent that intentionality really is something different from consciousness (and how different it is is debatable), we are puzzled as to how intentionality could give rise to consciousness. And saying that consciousness is recursivity, reflective self-awareness, a sense of self, or subjectivity is of no help because we have no idea what recursivity, reflective self-awareness, a sense of self, or subjectivity are. Indeed, explaining what they are has been a central challenge in the philosophy of mind for the last several decades.

The situation is actually worse than what we have portrayed. What really happens is that if *X* is proposed as an explanation for consciousness, then to the extent that *X* seems to capture consciousness, *X* is also very puzzling and not well-understood, so any sense of progress is an illusion. And to the extent that *X* is well-understood and not very puzzling, it is unclear how having *X* makes us conscious, so there is little or no sense of progress. It is a no-win situation.

Of course, it might be that only consciousness has all the suggested properties (and more?) at once, in which case the properties in question might be *jointly* sufficient for consciousness. In that case, localizing their intersection would be a helpful idea. However, claiming that they are jointly sufficient – that once you get all the relevant items together, then you get consciousness – requires an argument. Such an argument doesn't exist, yet.[3] And anyway, the move is desperate: what it really amounts to is making consciousness a grab-bag for all interesting or perplexing cognitive attributes. The hope seems to be that identifying consciousness with a smorgasbord of properties might actually

work: if we throw enough intellectual spice on our theory, we will cover up any residual deficiencies.

So, far from there being any agreed upon explanation of consciousness, we have a surfeit of explanations which range far and wide. It isn't often appreciated how troubling having all these explanations really is. The sheer plethora of proposals ought to be a warning sign to mind scientists of all stripes: It is clear evidence that something is very wrong.

Furthermore, having multitudinous explanations is *not* a sign that research proceeds apace. In fact, it means that research on consciousness is otiose. Having myriad explanations indicates that there are few or no scientific constraints forcing researchers in a particular direction. Without constraints, researchers can say pretty much what they want. A further consequence of this is that there is nothing forcing researchers to nail down a third-person account. Not only is no research direction indicated, but no research *at all* is indicated. Of course, research on consciousness is perhaps useful for its own sake, but it is clear that we do not need a theory of consciousness to conduct our other mind and brain sciences. For example, psychology and cognitive science proceed along happily without having even to mention consciousness (one can find such discussions, of course; our point is that they are not *required*). It simply doesn't matter for our experimental and theoretical work in these sciences whether we unite behind a theory of consciousness.[4]

There is an upside to not needing a theory of consciousness. It means that we can get on with our science of the mind. Given how hard it is, if we had to solve the problem of consciousness before we could do cognitive science, for example, then we wouldn't have any cognitive science.

It is our view that researchers have so little useful to say about consciousness because we are epistemically prevented from being able to connect consciousness to third-person processes. All researchers are really doing is pointlessly dousing each other with concoctions from their intuition pumps. We seek a different path.

1.2 The real debate: The naturalists and the mysterians

1.2.1 Defining naturalism and mysterianism

Researchers who study consciousness can be grouped into several camps depending on whether they believe consciousness is a physical (material) property. There are reductionists (or materialists) who think that consciousness will

both ontologically and epistemically reduce to (or is) some brain process. There are materialists who think that consciousness is a brain process, but are dubious we will ever understand or explain this fact. There are eliminativists who don't believe that there is any such thing as consciousness (a hopeless position, obviously). There are idealists who hold that consciousness is really all there is, that there really is no physical or material stuff or properties at all (this position, too, seems a tad drastic; ... still...). There are dualists who believe that consciousness is a nonmaterial property of the material brain. There are dualists who believe that consciousness is a nonmaterial property of some nonmaterial stuff in the universe, and that somehow this stuff gets associated with material brains. (These aren't all the positions there are, but these are the major ones.)

Cutting across all these camps is another division: those who believe that there will one day be a scientific theory robustly explaining consciousness, and those who believe that there will never be such a theory. This division is the one we are interested in.

Those who believe that science will one day develop a theory of consciousness that explains how it really works we call *naturalists*. And those who believe that there will never be any scientific theory of consciousness we call *mysterians*.[5] Naturalists believe that a satisfying explanation of consciousness is in the offing because they believe that a scientific theory of consciousness is in the offing, and they believe that theories offer satisfying explanations. Mysterians, on the other hand, believe that there will be no theory of consciousness because there can be no satisfying explanation of it, and, again, they believe that theories offer satisfying explanations.

We are neither mysterians nor naturalists. When we realized this, we realized that we *all* had been missing an important distinction: the distinction between a theory and a satisfying explanation. We hold that being a dualist, materialist, eliminativist, idealist, or whatever, is not all that important nor is it all that interesting. What really counts is what you think you can make of your position – specifically, whether you think a scientific explanation of consciousness will be forthcoming.[6]

Naturalists are not the guys in white hats nor knights in shining armor. They are not scientists poised on the brink of a great and deep insight. Naturalists believe that a scientific explanation of consciousness lies in our future. Some believe it lies in our near future (some even believe, incredibly, that they have such an explanation now, e.g., Dennett 1991). Such researchers have not been paying attention. As we pointed out above, that there are so many radically different theories of consciousness ought to give sober researchers pause. Clearly, consciousness is unlike anything we have theorized about previously.

Naturalists are wrong to believe that a satisfying explanation of consciousness is around the corner. But we take a stronger position. We think that a satisfying explanation of consciousness is not around *any* corner. A scientific *theory* of consciousness might lie in our future, but not a satisfying explanation.

On the flip side, we do not use the term "mysterian" in a derogatory way. Mysterians are not Luddites. Rather, they believe, for principled reasons, that a scientific theory of consciousness is impossible. They are wrong in this belief. Mysterians are, however, correct in believing that there will never a scientific *explanation* of consciousness.

1.2.2 The tension between naturalism and mysterianism

Most devout, committed naturalists believe that, though still poorly understood, consciousness is ultimately unmysterious or perhaps only mildly mysterious, a feeling which will fade with the time. These researchers have faith that science as it is presently construed will someday explain consciousness. One doesn't have to be a materialist to be a naturalist, though most are. There are dualists who believe that a scientific, naturalistic dualism is on the horizon – we just need to broaden our conception of causation, and probably our notion of explanation, by including in it notions of association or acquaintance or the like (it is not clear this broadening will work, since on all dualisms, the kind of tight conceptual connection that makes our best scientific theories so robust and satisfying is unavailable, no matter what notions one adds; one can therefore legitimately wonder if "scientific dualists" have changed the game so much that they are no longer contributing to science, but rather, to speculative philosophy or metaphysics (see, e.g., Chalmers 1996, 2003)). One could even be an idealist and be a naturalist: one would just have to believe that some of one's experiences will explain one's other experiences scientifically. We don't know of any idealist naturalists, so if you are looking for some territory to stake out, this one is available. The important unifying conviction for all naturalists, though, is that we humans will be able to understand consciousness in some sort of reasonable and satisfying third-person way, through some sort of rigorous process of scientific inquiry.

On the other side are the mysterians. These researchers are not sanguine at all about explaining consciousness, scientifically or otherwise. They all believe that whatever consciousness is, it is truly and deeply mysterious and will remain that way. And, as a result, science will have little to say that is truly informative about consciousness. In particular, science will never be able to explain why brains are things that are conscious, what parts of the brain produce consciousness, why some things have inner lives and others don't, nor will it ever explain why, e.g., trumpets sound like they do to us, nor why chocolate ice cream tastes the way it does. They believe that consciousness is almost too bizarre to be real. They take that lesson very seriously and despair.

As with the naturalists, mysterians can either be materialists, dualists, or idealists. So, for example, there are materialists who believe that though consciousness is certainly a material property of the brain, science will never succeed in understanding it. Some materialists who believe that science will never understand consciousness actually hold a more complicated view. They believe that though *human* science will never understand consciousness, a radically different kind of organism might in fact have little trouble developing a theory of consciousness (e.g., see McGinn 1989, 1993). These materialists don't believe that consciousness, as such, is a mystery, i.e., it is not *fundamentally* a mystery. Rather, it is mysterious *to us*, and to other creatures like us. Other materialist mysterians, however, do, in fact, believe that consciousness is fundamentally mysterious – how it arises will be a mystery to any creature who contemplates it. And, accordingly, no scientific theory of consciousness is in the cards (Dietrich & Gillies 2001).[7]

These naturalist and mysterian camps cannot talk to one another, for their differences are deep and entrenched. Some see the conflict as over whether or not consciousness is a brute fact about the world. Others see it as turning on whether consciousness has any relevant causal properties. However, in large part, the divergent reactions between the naturalists and the mysterians depend on antecedent views about what counts as an explanation in science. The naturalists are those who are sold on the promise of science. They believe that the way to explain something is to build a model of it that captures at least some of its etiologic history and some of its causal powers. Their approach to explaining consciousness is to isolate the causal influences with respect to consciousness and then model them (see, e.g., Churchland 1984; Flanagan 1992; Hardin 1988).

In contrast, mysterians do not believe that science and its commitment to modeling causal interactions and organization are always up to the challenge of explaining the world via theory, and this is especially true when it comes to providing a theory of the conscious mind (e.g., McDowell 1994; Nagel 1979; and, perhaps, Block 1997; Searle 1992). They believe that some things – many things – might be scientifically explained in terms of physical causes, but qualia aren't going to be one of them. Isolating the causal relations associated with conscious phenomena would simply miss the boat, for there is no way that doing that will ever capture the qualitative aspects of awareness. What the nat-

uralists might do is illustrate *when* we are conscious, but that won't explain *why* we are consciousness. The naturalists would not have explained why it is that 40 Hz neural oscillations, or the activation of episodic memory, or an executive processor, or whatever, should have a qualitative aspect, and until they do that, they cannot claim to have done anything particularly explanatory with respect to consciousness.

Another way to couch the difference between naturalists and mysterians is in terms of when each camp thinks the enterprise of explaining consciousness will be completed. Mysterians say that the project of explaining consciousness will not be finished until the existence and nature of qualia are explained in some satisfying, third-person, scientific terms. Naturalists usually agree with this. But there are four main varieties. Some naturalists take the just-wait-until-next-year attitude: "We will explain qualia completely and without remainder soon – please be patient." Another group says that explaining the relevant neural or cognitive states associated with consciousness *is* explaining qualia. A third group significantly weakens what counts as an explanation. They hold that explanations of consciousness will flow solely from laws that associate or correlate the phenomenal realm with the material or physical realm. When such correlations are listed, nothing of any interest will remain to be explained, or at least nothing of any *scientific* interest will remain to be explained. Members of all three groups believe that the explanations they foresee will flow from some theory, and will be satisfying, at least to a fair degree. Members of the first group refreshingly stick to familiar notions of a satisfactory explanation. That is why they put all their bets on the future. Members of the last two groups do not place all their bets on the future; they try to manipulate or try to change the notion of a satisfactory explanation, often by delving into the psychology of what makes an explanation satisfactory. Some of these naturalists think that the satisfaction supplied by their eventual theory might be an acquired taste – a state achieved only after living with the new theory for quite sometime: Familiarity engenders explanatory satisfaction. Given the theory and enough time spent using it, scientists will eventually lose any sense that consciousness is mysterious and the theory's explanatory capacity will be found to be quite robust. Lastly, there are some wily naturalists who admit that the correct explanation will certainly *not* seem true or strike us as true, but that we will be able to discount this feeling because of the power of the theory (McDermott, 2001, tries this move; for our reply, see Dietrich & Hardcastle 2002).

Mysterians, not surprisingly, have something to say to each group. To the first group they say: "Yeah, yeah." To the second group they say: "You are not explaining consciousness, but something weaker. And you cannot dodge this fact merely by stipulation." And to the third group they say: "This kind of explanation is so weak, it won't be remotely satisfying. The conceptual framework which includes consciousness and the one which includes the rest of the universe supposedly tied to it via 'psychophysical laws' are too incompatible to provide any sort of satisfying explanation." And to any naturalist who tries to alter the notion of satisfactory explanation or to explain away the feelings of unsatisfactoriness, they say: "Satisfying explanations are not merely an acquired taste. They encapsulate real, robust logical connections. All that spending time with a theory engenders is skill with the theory, but no one who is honest confuses skill with explanatory satisfaction." And then to all four groups the mysterians say, "Since none of you has a genuine explanation of consciousness, none of you will be able provide the world with a theory of consciousness."

1.2.3 Silence at the impasse

Here is a summary, with our commentary, of how the argument has gone between the naturalists and the mysterians.

The naturalists begin: "Let us assume a prior and fundamental commitment to mechanism. If we are materialists, then we have to believe that consciousness is some physical mechanism, presumably something in the brain. So we ought to be able to isolate the components of the brain and of brain activity necessary and sufficient for consciousness." (The story is more complicated if the naturalists are dualists, but the basic outline remains the same: there is a commitment to mechanism and from that they get some necessary and sufficient physical *correlates* for conscious states. These states would then *index* the structure of consciousness in all its glory, and such correlates would function in some sort of bridge laws that would form the basis of the science of consciousness.)

The mysterians reply: "Though the naturalists might have been successful in isolating the causal etiology of consciousness, they have not explained why it is that that isolated brain activity should result in conscious. They have not explained why it is like anything at all to have that brain activity."

"Part of a *good* explanation," the mysterians maintain, "is making the identity statement, or correlation statement, intelligible, plausible, reasonable, intuitive. This is a conceptual point: the explanans doesn't need to make the explanandum necessary, but there should be some type of inferential relationship – a plausible, probabilistic, inductive one is perfectly acceptable – between the concepts of the explanans and the explanandum. So, the concept of con-

sciousness ought to follow or arise naturally and easily from our conceptions of brain functioning. The naturalists have not done this: the facts surrounding consciousness, and whatever facts in the material realm are alleged to explain it, strike most researchers as light-years apart. Why this should be is very interesting. The concepts we use to think about consciousness and those we use to think about neurons, neurochemistry, psychology, cognition, and the information-processing mind seem unrelated. Perhaps they are in fact unrelateable. So, naturalists have not explained the most basic, most puzzling, most difficult question of consciousness. They haven't removed the mystery of the connection between the conscious mind and its body." (Sometimes this unclosed connection is referred to as the *explanatory gap* (Levine 1983).)

The mysterians are, of course, right: being a good explanation *is* an explanans making the explanandum intelligible, plausible, reasonable, and intuitive. Of course, intuitions can be beefed up, exercised, and even altered, but unless some explanation appeals to our (perhaps tutored) intuitions and strikes us as plausible, it won't be satisfying. So, scientific theories of consciousness won't explain the weirdness of consciousness to those who find it weird. And, as we have argued above, we should all find it weird.

In response, the naturalists point out that at least some part of a theory becoming a satisfying explanation is just getting comfortable with the theory. They point out that old concepts have to change or die off. Then the new concepts supplied by the theory can work their magic and provide everyone with an intuitively plausible explanation. In this vein, naturalists parade other cases of scientific theories that at first seemed unsatisfying as explanations, but then later grew on us. For example, they say, "Like a 'life-mysterian' – someone who believes that life is something over and above biology and biochemistry – consciousness-mysterians need to alter their concepts. To put it bluntly: mysterians' failure to see that consciousness can be explained by a theory cuts no ice with science. Their concepts are at fault, not science."

But this ploy doesn't work; here the mysterians beat the naturalists at their own game. The mysterians say, "Life-mysterians held out for the view that life was something over and above biochemistry. They wanted but failed to get the missing 'life-force.' We, on the other hand, are *not* (or are not *necessarily*) holding out for the view that consciousness is something over and above neurochemistry. We simply deny that neurochemistry, or any other science, is going to explain consciousness satisfactorily. Just look at that sunset, taste this strawberry, listen to the thunder ... Where are these *experiences* in your neuroscience? Experiences might well be material, but explaining them scientifically is not in the cards. Life-mysterianism isn't reasonable; Consciousness-mysterianism is."

At this point, the naturalists make a very reasonable move: they point out not that today we can see that consciousness-mysterianism is like life-mysterianism, but only that we can't know today what science will discover in the future, that vast expanse of time wherein lives all hope – and science is nothing without hope. So, claim the naturalists, one day in the future, we might very well discover in some new process the very concepts we need to understand consciousness truly and satisfyingly. It is arrogant, say the naturalists, for the mysterians to claim that human scientists will *never* come up with a scientific theory of consciousness that explains this unusual phenomenon to us.

Here, the naturalists are on to something. A gauntlet has been thrown down: if you want to remain a mysterian, it is not enough to wring your hands and say, "Woe are we." You have to come up with an *argument* that an explanation of consciousness is not in the cards. Of course, several mysterians have tried to do just that (see, e.g., Levine 1983; McGinn 1989).

Finally, there is the following self-defeating move we have seen some naturalists make. They try claiming that it is just a brute fact about the world that consciousness is just such and such brain activity (or, for the naturalistic dualists, consciousness just is tightly coupled with or associated with such and such a brain activity). This is just the way our universe works and we should accept it without whimpering. This move is self-defeating because, by definition, it can never produce a satisfying explanation. Brute facts are brute, after all. Nothing explains them. So, for the naturalists to take this tack is to give up naturalism.

There is something else to say about this "brute fact" approach. Though at times it appears that the brute fact approach is what the naturalists are assuming, especially when they dismiss out of hand those overcome by the eeriness of consciousness, in an ironic meeting of the minds, this is what some mysterians want to do as well. Consciousness is too odd for us to grasp, so we should just accept it without expecting scientists or anyone else to plumb its mysteries. No further explanation is needed or expected. (One can read McDermott 2001, this way.)

However, this sort of response is too facile, on both sides. It is true that we accept brute facts about our universe. Our universe contains matter and energy. We don't spend much time wondering why, it just does, and we reason from there. On the other hand, there are facts about the world that we do not accept as brute. We feel perfectly comfortable expecting an explanation for why water is liquid. That is not a brute fact. We explain the liquidity of water by appealing to other facts about the world, the molecular structure of water

and its concomitant microphysical properties, for example. And these facts are explained in turn by other facts, such as the quantum mechanical structure and properties of hydrogen and oxygen. Now *these* might be brute facts, but perhaps not. Eventually, however, we will work our way down to some basic matter and its energetic interactions. This matter and its interactions explain not only the liquidity of water, but the physical nature of everything else, solid or gas, as well. Only a few, very privileged, and very fundamental facts about our universe are brute, and they underlie everything else in our world.

We are not advocating this rosy view of science here. We are merely pointing out the quite reasonable desire to have as few brute facts as possible. The way to achieve this is to make sure the brute facts are very basic, where this means that when working together, the small set of brute facts account for as much as possible. At this stage of inquiry, it seems that consciousness should not be classified as a brute fact (in general, whether some fact is brute or not is partly empirical and partly theory-based). It seems like consciousness should be due to more fundamental facts in the universe – everything else about the brain is, why should we expect consciousness to be any different? Hence, if one is to claim that consciousness being a 40 Hz oscillation or executive memory is simply a brute fact about the universe, then one is prima facie operating with an odd and perhaps troubling metaphysics. Of course, such a metaphysics might turn out to be true: consciousness might be a further brute fact about the universe (Chalmers's argues for this). Nevertheless, we don't want to claim consciousness is brute unless we absolutely have to.

We have reached a standoff. The naturalists say materialism, or dualism and some sort of bridging mechanism, entail an identity statement, or bridge law, for consciousness, and that such an identity or bridge law will be intelligible, plausible, reasonable, and intuitive. Consciousness will eventually be no more mysterious to them than the liquidity of water or the aliveness of living things. A robust explanatory theory should someday come. The mysterians do not share these intuitions. Whatever identity statements the naturalists dream up simply won't be enough to persuade them that consciousness has been explained in any theoretically interesting way. Understanding consciousness requires more – much more – than even perfect correlations. And since no explanation is in the offing, there will be no theory, either.

It is obvious that the naturalists and mysterians agree about the facts: We don't have a scientific theory of consciousness, but one day we will be able to get a series of relevant correlations – maybe. But their reactions to these facts are quite different. The naturalists see the (hoped for) correlations as a promise

for a future theory and robust explanation. The mysterians see the possibility of getting correlations as a dead-end, and hence as a dismal defeat for science.

Our position about the disagreement between the mysterians and the naturalists is that they are both wrong. They are both wrong because both the mysterians and the naturalists have the same understanding of "theory;" they both agree that scientific theories offer explanations. Sometimes this is true. We certainly like it when it is true. But explanations are not necessary adjuncts of theories, and in the case of consciousness (and some other interesting cases), theory and explanation pull apart. It is an interesting story how this comes about and the consequences it has for our understanding consciousness (and other things as well). But first, we need to explain why the naturalists and the mysterians are both wrong.

Part I

Troubles with naturalism

Chapter 2

Against naturalism

The logical boundary of conscious perception

> We know two things about what we call our psyche (or mental life): firstly, its bodily organs and scene of action, the brain (or nervous system) and, on the other hand, our acts of consciousness, which are immediate data and cannot be further explained by any sort of description. Everything that lies in between is unknown to us, and the data do not include any relation between these two terminal points of our knowledge. If it existed, it would at the most afford an exact localization of the processes of consciousness and would give us no further help toward understanding them.
>
> *Sigmund Freud*

There is an intuition most of us have – whether we are mysterians or naturalists – that our conscious experiences could somehow be sundered from the world experienced. This intuition usually manifests itself in two forms: the *Cartesian* and the *zombie intuitions*. These two intuitions are typically used to generate metaphysical claims – that is how dualists use them, for example. But in this chapter, we are going to discuss their *epistemic* force: these intuitions have the power to derail naturalism – a fact not generally appreciated. We also present the best way to deal with them from the naturalist's point of view. But then we show that that way requires something logically impossible and so is untenable. Our conclusion will be that the two intuitions are *ineluctable*: once held or once acquired, there is no epistemic condition that can remove them. It follows from this that naturalism, of any stripe, is not a viable option: there will be no explanatorily satisfying theory of consciousness.

Before we turn to that, however, we briefly explain the technical notion of *supervenience* (Chalmers 1996; Davidson 1970; Kim 1993), and discuss how we might come to believe that a supervenience relation exists between two things.

2.1 Supervenience and its epistemology

Supervenience is crucial to the modern consciousness researcher. It is a notion usually couched in terms of properties: properties supervene on other properties. In general, *A* properties supervene on *B* properties when fixing the *B* properties automatically fixes the *A* properties. There are two ways in which *A* properties can supervene on *B* properties: naturally or logically. Natural, or nomic supervenience, refers to items only in this universe, constrained by our natural laws. *A* properties supervene naturally on *B* properties if, according to the laws and conditions of this universe, any two situations identical in their *B* properties are identical in their *A* properties. Logical supervenience refers to any logically possible world (any logically possible universe). *A* properties logically supervene on *B* properties, if any two logically possible situations identical in their *B* properties are also identical in their *A* properties. In both cases, *B* is called the *supervenience base.* (Another way to state logical supervenience is to say that *A* facts supervene on *B* facts if fixing the *B* facts suffices for fixing the *A* facts. We will move between these two characterizations without much ado. Nothing turns on this, for a fact is a metaphysical particular pairing some object and some property. It is a fact that 2 is prime, and it is a property of 2 that it is prime. The fact pairs the number 2 with the property of being prime.)

It is important to understand logical supervenience thoroughly, so we consider some examples. Imagine a glass of liquid water. The molecules in the glass are caroming all over the place in an agitated way. Now, try to imagine another glass of water where the molecules are behaving in exactly the same way as in the first glass, and all other relevant microphysical facts are the same, but where the water in the second glass is frozen solid. You can't do it (you are mistaken if you think you can). The physical state of the water, solid, liquid, or gas, is determined by the motion of the molecules: change the motion significantly, and you change the overall state of the water.

This observation can be extended to any temperature of the water in the glass, whatsoever. Imagine a glass of hot water, say 95 degrees Celsius. The water molecules in such a glass are very agitated. It is logically impossible that another glass of water could contain water molecules behaving exactly as in the first glass, and yet the water be at, say, 20 degrees Celsius. Temperature *is* molecular agitation. This is a conceptual truth, arising partly from how temperature is defined, and partly from our understanding of the physical world. Fix the behavior of the water molecules in the glass and you automatically fix the water's temperature. (It is important here to focus on temperature and not how hot or cold the water feels since how it feels is a phenomenological state.)

Another good example of supervenience is in your computer. Fix the states (1 or 0) of all the bits in your computer (in the CPU, on all the RAM, and on your hard drive, etc.), and you thereby establish the program you are running (a word processor, say), what you are doing with it (writing a chapter, say), and what parts and versions of the chapter are stored on your hard drive. The bits are the low-level facts, and the chapter as it appears on the screen as you write it is the high-level fact. Fix the bits, and you fix the chapter in its entirety. Change even one bit, and you change something, e.g., the document or the program.

Our physical universe as a whole works just like the computer or the glass of water. It is simply impossible that the low-level facts about our world could be exactly what they are and yet there be no stardust, no suns, no galaxies, no planets, no continents, no minerals, no life, no US Constitution, no penguins in Antarctica, and no MTV. In short, and though it may sound strange, MTV is what it is because certain low-level facts are what they are. There is no possible world with all the same, subatomic facts as ours that isn't blessed with MTV.

Since fixing all the low level facts in our physical universe completely determines what all the high level facts are (note, this does not go the other way), in an important sense our universe is just one big set of physical facts described at different levels. The lowest level is the subatomic level (or some sub-subatomic level), from there we move up through the atomic level, the molecular level, the macrochemical level, the biological level, and on up to the level of cultures and politics.

To understand that a supervenience relation actually exists between some low-level facts and some high-level ones, one makes an inference. This inference connects the low-level facts with high-level ones in one's mind. Only after this inference has been made can one see that the relevant properties supervene on their base. We call this inference a *supervenience inference.*

Supervenience inferences can be very well-behaved sometimes. This happens in cases of low-complexity. If some molecules form themselves into a square then squareness supervenes. One can just "see" in one's mind's eye that were some molecules to do this, squareness would obtain. This is an easy conceptual matter. But much more common are the complex causal inferences. The motion of a car is causally dependent on the motion of its engine and drive train: the motion of the car supervenes on the motion of its engine and drive train, for fixing the motion of the drive train suffices to fix the motion of the car. The medievals called this kind of causation "causation *per se*." They called the kind of causation where, for example, a rolling bowling ball knocks over some bowling pins "causation *per accidens*." What distinguishes causation *per se* from causation *per accidens* is that, in the former, if the effect exists,

the cause must too (Priest 2002: 35). We stress that causation *per se* always requires fixing the context of causation. So, for example, in the car case, it must be assumed that the tires are on a road, that the road has the appropriate friction, that the coefficient of friction is correct, that the air pressure is standard, etc. etc. When talking about causation *per se* we will always assume that the context is fixed appropriately without specifying what that might entail. When discussing supervenience and causation, we will always mean causation *per se*.

Seeing some specific case of supervenience as causation *per se* requires making at least a somewhat complex causal inference. In principle, this inference is like seeing the supervening of squareness, but in practice many more concepts and conceptual relations are involved.

2.2 Zombie and Cartesian intuitions: Roadblocks to an explanatory theory of consciousness

2.2.1 The nature of the zombie and Cartesian intuitions

One version of the intuition that consciousness experience can be sundered from the world experienced or the body having the experiences is the intuition that consciousness is completely unnecessary for thought and action; consciousness just comes along for the ride.

This view is highlighted in the notion that consciousness *is* experience: things happen to us and we experience them. All the real work of the mind is done by cognitive, subcognitive, and noncognitive processes. But the felt disconnect is deeper than this. Not only is the real work done by mental or neural processes that don't need to be conscious, but consciousness is not even adequately causally or logically connected to the mind. If true, this intuition suggests that there might be creatures who lack consciousness entirely but who otherwise behaved perfectly normally, or even perfectly rationally. Such creatures are called *zombies*. The intuition that such creatures are possible is called the *zombie intuition*.

The zombie intuition develops in virtually anyone who philosophically wrestles with the nature of consciousness. (We say that one needs to wrestle philosophically with the nature of consciousness because it is far from clear that laypeople have the zombie intuition. However, we are confident that we could instill or activate this intuition in most laypeople given an afternoon; our students are relevant, positive data.) That one can sunder the phenomenal and the physical, and implement just the physical in a non-phenomenal system

just seems to be obvious, if not immediately obvious. And anyone who has the zombie intuition is going to find it difficult to conceive of consciousness as logically supervening on the physical – even if it does and even if they were shown the neurally relevant supervenience base. Yet, any decent naturalism is going to have to somehow sideline or explain away the zombie intuition.

There is another intuition in the area that is much more common that has the same results. Many laypeople develop this intuition easily and naturally on their own. This is the intuition that our conscious experiences could be just what they are regardless of how the world is – that somehow our consciousness need not cohere with how the physical world actually is. We call this intuition our *Cartesian intuition*. Like the zombie intuition, it, too, is a kind of belief (as are all intuitions).

Descartes obviously had this intuition (see, for example, Part IV of his *Discourse on Methods* and other places). Interestingly, it is reasonably clear that Descartes didn't have the zombie intuition: he thought zombies were (morally) impossible. In *Discourse on Methods*, Part V, he says:

> . . . if there were machines bearing the image of our bodies, and capable of imitating our actions as far as it is morally possible, there would still remain two most certain tests whereby to know that they were not therefore really men.

The two tests are language (Descartes was certain that zombies (or machines) would never be able to truly use a natural language as humans do) and the ability to act based on subtleties of reason and knowledge (Descartes thought that, eventually, a zombie would betray itself by not being able to do something which involves reason that humans can readily do).

The Cartesian intuition (as Descartes noted) is easy to come by. Dreams are a good route. At one time or another, we have all dreamed that we are somewhere strange or have vividly imagined that we are doing something exciting or novel, yet we are not where we dreamed nor doing what we imagined. If we can dream we are hang gliding in Tibet when we are home in our beds, then perhaps we are just dreaming we are at home in our beds. Perhaps everything is a dream. Perhaps boarding a jet, flying to Nepal, trekking in to Tibet, and hang gliding is a dream. We all have at one time or another thought something like, "What if none of my experiences are real? What if my entire experience of the world is one big dream? What if nothing is the way it appears to me; what if nothing is the way I experience it?" It can be a frightening thought. But it could be true, it seems.[1]

The Cartesian intuition is another bar to naturalism. Like the zombie intuition, this intuition sunders the physical from the phenomenal. But it does it differently than the zombie intuition.

The zombie intuition makes a claim about supervenience. It says that consciousness doesn't logically supervene on neural processes in the brain: consciousness needn't exist just because the relevant neural prodsses do. The world is there, but it is not experienced. The Cartesian intuition does not make any supervenience claims. To say that *A* supervenes on *B* is to say all *B* worlds are *A* worlds. It is *not* to say that a C world couldn't also be an *A* world.

Couched as claims about how the world is experienced, we can render the zombie intuition as:

Different minds / Same world

and the Cartesian intuition as:

Same minds / Different worlds.[2]

The zombie intuition says that minds can vary while the world remains the same, and the Cartesian intuition says that the world can vary while minds remain the same. Either way, the result is the same: there is no connection between minds and world, between conscious experiences and the alleged causes of those experiences strong enough to support an explanatory science of consciousness. Both intuitions make it mysterious why certain brain states result in or cause the phenomenal experiences that they do. The zombie intuition does this by blocking the relevant supervenience inference. The case of the Cartesian intuition is more involved. Here, two minds having the same experiences could reside in radically different worlds. There might be supervenience inferences available from the worlds to the minds, but there would be no explanations because nothing would explain the given phenomenal states (the experiences) as a *type*. Explanations require more than supervenience relations. Suppose we discover two containers of water. In one, the water is frozen and in the other, the water is boiling. But both are at a temperature of 50 degrees Celsius. Assuming the case was repeatable, and assuming we could convince ourselves that our thermometer was working (and everything else was working in standard fashion – both are large assumptions), we would have little choice to but to revise thermodynamics.[3]

If one seeks a scientific explanation of consciousness, then the Cartesian and zombie intuitions are serious obstacles. A good, satisfying explanation of consciousness will need to make the connection between it and the brain/neural states on which it supervenes more than just intelligible, though

it will need to do that, too. The explanation will have to make the connection *compelling*. If the explanation cannot do this, then there will be little reason to call it an explanation as opposed to just a statistical association. The term "explanation" might be used, but it would be honorific. The intuitions, especially the Cartesian intuition, which is so easy to come by and so easily held, will make any proposed explanation seem thin, even to specialists who will be well-acquainted with technical details of the theory. Both intuitions drive a wedge between the very concepts that need connecting and that must be connected if an explanation of consciousness is to work, if the explanation is to be worthy of the name. Therefore, some way has to be found to sideline these two intuitions.

2.2.2 Persistent illusions: The best case for naturalism

Scientific explanations often show us an aspect of the world different from our intuitive understanding of the world. In general, any proposed reductive explanation, if it is to be satisfying and compelling, has to explain away, or reduce the force or the compellingness of these countervailing intuitions. We refer to this reduction of compellingness as *defanging* the countervailing intuitions.

Defanging intuitions that stand in the way of a scientific theory can be accomplished in a variety of ways. If the intuitions are veridical, then they must represent knowledge of some genuine physical process. In that case, the best way to handle them is to explain them as a natural consequence of the other physical processes, or, in the best case, as natural consequences of other processes working in tandem with the main physical process in question. Here is an example of what we mean.

Galileo's theory of gravity had the consequence that a hammer and a feather, if released at the same time and from the same height above the ground, will hit the ground at the same time. Of course, if one runs this experiment in the open air on Earth, the results are quite different from the predicted results – the feather comes to rest on the ground much later than the hammer. Is Galileo's theory wrong? No. Air resistance is causing the problem: the feather has much greater air resistance than the hammer and so is slowed down much more by the air. Galileo's theory assumes an airless environment. Run the experiment on the Moon, or in an airless container here on Earth, and you get the correct results. The plausibility and hence believability of Galileo's theory goes up with the introduction of air resistance to explain why on Earth, the feather floats down long after the hammer. And of course, gravity explains why there is air and air resistance, in the first place, so the theory just gets stronger.

But though common and very useful, this sort of defanging cannot be used in the case of consciousness. This is because the countervailing intuitions in the gravity example and in the example of consciousness have different epistemic status owing to their different etiologies. On Earth, the feather really does reach the ground long after the hammer. The intuition that gravity moves heavy objects faster results from observing such facts. But the situation with consciousness is nothing like this. We don't, and indeed cannot, observe zombies, and no one can conclusively prove via other means that zombies are possible (we merely think they are possible (if we do) because they are conceivable). On the other hand, it is impossible to prove that there aren't zombies. So we don't know if there really are zombies or not. The relevant observations– basically that there is an epistemic gap – are, at best, inconclusive. The best strategy for the naturalist then, is to construe the Cartesian and zombie intuitions as *persistent illusions*. This view of them says that the zombie and Cartesian intuitions are not knowledge. They are, therefore, not suggestive of nor guides to the nature of reality at all. They are misleading appearances, but appearances of such power that we can't shake them. The argument for this is made by considering these exhaustive three cases.

Case 1: Zombie twins

The crucial intuition used by Chalmers in his argument for dualism is the intuition that we each have a logically possible zombie twin (1996). Your zombie twin is your exact physical duplicate in another possible world, but where this duplicate lacks your conscious experiences. In the appendix to this chapter, we show that the argument for zombie twins does not and cannot be made to work. And we argue for something stronger, that the zombie twin intuition is incorrect: no conscious person or creature of any sort could have a zombie twin. But understanding this argument has no effect on dualists, nor even on us who accept it – we ourselves continue to have *prima facie* beliefs that we could have zombie twins. (A *prima facie* belief is a defeasible, easily held, but not deeply considered belief. Perceptual beliefs are good examples of *prima facie* beliefs.) This shows that the zombie twin intuition persists even in the face of demonstrations of the impossibility of zombie twins. Hence, the zombie twin intuition must be an illusion – a deeply persistent illusion.[4]

Case 2: Non-twin zombies

Non-twin zombies are possible, perhaps. These are rational, cognitively robust creatures in some possible world who nevertheless lack consciousness. They are not clones of us, but they might have great civilizations, etc., while being as conscious as doorknobs. Many philosophers can easily imagine such creatures. But imagination is doing all the work here. No one ever actually observes such zombies. Even if you are the only conscious creature in the entire universe and all the rest of us are in fact zombies (a version of the problem of other minds), you still don't observe us *qua* zombies. Again, this case is completely unlike the Gallilean case. That we can readily imagine non-twin zombies is not particularly good evidence for, nor a particularly good reason to believe the claim that non-twin zombies are possible. Hence, that we can imagine non-twin zombies is best thought of as an illusion. That is, lacking good evidence that non-twin zombies are possible, the best epistemic status that the notion of non-twin zombies can achieve is that of being a persistent illusion.

Case 3: A Cartesian world

In the case of our Cartesian intuitions, we really do observe something relevant to it: every night we dream and hence consciously experience events that didn't happen. This is evidence for our Cartesian intuition. But, it is only mild evidence. Humans' Cartesian intuitions vastly outstrip the paltry evidence supplied by dreams and the like. We have no evidence that the entire world, in all of its rugged coherence, could be a sustained lie. That is a gigantic generalization of the mere notion that sometimes we experience things that don't exist. As an antidote to this, notice that the large majority of us have no trouble at all distinguishing between dreams and reality (i.e., the non-dreamed world). Our Cartesian intuition is quite robust, and is quite unlike the Galilean case. So, again, this intuition is best construed as a persistent illusion.

Consider the Müller-Lyer illusion (Figure 1). The two horizontal bars are exactly the same length (measure them, if you wish). We know that they are the same length because in making this illusion we first drew one horizontal bar and then, using a drawing program, duplicated it to produce the second bar. They are "clones" of another. Yet so compelling is the illusion that even we had to re-measure them to make sure we hadn't made a mistake. The Müller-Lyer il-

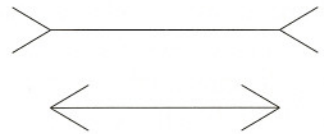


Figure 1. The Müller-Lyer illusion. The two horizontal bars are exactly the same length, yet the top bar looks longer

lusion is so persistent that no amount of arguing or measuring or even cajoling will be enough to alter one's perception that the top bar is longer. Psychology is rife with such illusions. They persist even in the face of complete refutation. The zombie and Cartesian intuitions work like this. At least, so should say the naturalist.

Naturalism's best chance for success at defanging the Cartesian and zombie intuitions, then, is to treat them as illusions like the Müller-Lyer illusion. We mean this analogy quite precisely. We are saying that Cartesian and zombie intuitions are *perceptual-cognitive illusions* just as is the Müller-Lyer illusion a perceptual-cognitive illusion. Indeed, that it is perceptual is the going explanation of why the Müller-Lyer effect persists even in the face of rational refutation: the perceiving, so the explanation goes, is under the control of a module that is informationally encapsulated from higher cognitive processes like rational decision-making and the like (see, e.g., Fodor 1983). We aren't committed to the modular explanation being correct, in either the Müller-Lyer case or in the case of the consciousness illusions. All we need is that the Cartesian and zombie intuitions are illusions very much like the Müller-Lyer illusion, and that these illusions persist because of their perceptual-cognitive nature.

Given this, the following analysis of the situation emerges. First, note how one goes about counteracting the Müller-Lyer illusion: one measures the two horizontal bars and discerns that they are the same length. This evidence is given a greater weight than the other perceptual evidence confronting one's eyes. One ignores then what one's eyes are telling one and believes (perhaps with an act of will) that the two horizontal lines are the same length. It is important to note that measuring the horizontal bars results in a different kind of perceptual information, not in some nonperceptual information that is somehow better than the perceptual information one gets by just looking with one's eyes. The measurement-based perceptual information is given a greater weight because of its etiology, not because it is a different kind of information.

Just so with the supposed eventual theory of consciousness imagined by the naturalists. Eventually, the naturalists claim, someone will produce the correct explanatory, reductive explanation. This explanation and its concomitant theory will provide some sort of robust evidence that such and such a brain process is the seat of our consciousness (the supervenience base of our consciousness). Given the compellingness of this evidence, we will naturally weight it over the Cartesian and zombie intuitions and come to believe what the theory says is true of consciousness rather than what our intuitions tell us is true. The two intuitions won't go away in the slightest, but they will be rendered impotent – defanged – because they will be seen to be pure illusions. It is in this

way that the eventual correct theory of consciousness will win our hearts and minds, if not our intuitions (McDermott 2001, makes precisely this point when defending this theory of consciousness; but see Dietrich & Hardcastle 2002).

What would make the evidence compelling? It is clear that mere statistical data correlating some neural processes with certain conscious states are never going to be enough to sweep aside the zombie and Cartesian intuitions, rendering them mere persistent illusions. Perhaps there are other ways that evidence from a naturalistic theory could compel us to believe the theory, but we don't know what they are. There is, however, one obvious, best case: just as in the case of the Müller-Lyer illusion, perceptual information with an impeccable etiology would be compelling to all concerned – to naturalists and mysterians alike. The indisputably best perceptual information would be to witness consciousness arising from its supervenience base. This evidence obviously would outweigh the zombie and Cartesian intuitions and sweep all doubt away (though not the intuitions). Materialism would be established, naturalism would therefore be correct, and we would be in possession of the long sought-after theory and concomitant satisfying, reductive explanation of consciousness.

Alas.

2.3 The logically hermetic nature of consciousness

2.3.1 The hermetic property

Consider the question: "What would you experience if you saw the process on which consciousness does logically supervene?" (If it helps, assume that materialism is true.) Answer: You would just experience another quale. So you couldn't, by definition, see that quale cause or result in your conscious experience because in experiencing the processes as some quale, you are already conscious – that's what it means to experience a quale. Put another way, you can't see the supervenience relation between your consciousness and your consciousness's supervenience base because you can't step outside your conscious experience.

This argument needs more detail. Assume that our consciousness does logically supervene on some neurological process, the so-called neural correlate of consciousness (NCC).[5] The NCC is the supervenience base of consciousness, let us suppose. The goal now is to witness, literally see, the supervenience relation in action. Is this possible? No. For, what would you see if you saw your

NCC result in your consciousness?[6] You would experience just another quale: you would see (visually experience) some working neurons, for example. That is not seeing or experiencing the logical supervenience relation. It couldn't be, because experiencing the NCC is not experiencing the supervenience base as it must be experienced to satisfy the goal – it is not experiencing the NCC *as* the supervenience base. Experiencing the NCC as a quale is not experiencing the NCC as giving rise to consciousness; experiencing the NCC is merely having yet another conscious experience. Experiencing your NCC presupposes that you are antecedently conscious, hence your NCC has already done its job, hence you can't see it do its job.

One can, of course, experience that which results in one's experience: one can see the brain processes on which one's consciousness supervenes, the NCC. But you cannot experience the actual supervening relation – you cannot experience your NCC causing *per se* your experience of the NCC.

We are encapsulated inside our consciousness. We can't see anything causing it or resulting in it because we can't step outside our consciousness and watch it come into existence via some causal process. (And, of course, we can't see it happening in anyone else either.) Our claim is that for us to understand how our consciousness logically supervenes on our NCC (assuming that it does), we would have to see the NCC actually result in some consciousness, which is precisely what we can't do – as a matter of *logic*.

Searle's famous Chinese-room thought experiment (1980) provides another argument for this conclusion that directly seeing the supervenience relation is logically impossible (we assume familiarity with the thought experiment). Assume, contra Searle, that the room is conscious and that the room's consciousness logically supervenes on what the person-in-the-room does (this assumption is similar to what Searle called the System's Reply). Also, assume that the room is made of clear Plexiglas, and that the room is looking in a mirror. Finally, assume that the room's capacity to see is also due to what the person-in-the-room is doing.

The room qua consciously experiencing entity cannot see its consciousness logically supervene on what the person-in-the-room is doing. The room can see in the mirror what the person-in-the-room is doing, to be sure, but any experiences the room has (which by assumption are due to the actions of the person-in-the-room and which supervene on these actions), including experiencing seeing the person-in-the-room, are just that – *experiences*. The room can't see those experiential states result in its conscious experience because those states are not part of the supervenience *base*; they are *supervening* on the supervenience base – they are the conscious experiences themselves.

If the person-in-the-room stopped her activity, then the system would stop being conscious. If the person started up again, then consciousness in the room would resume. But the room wouldn't be able to see, to perceptually experience, the transition of the person-in-the-room from no activity to activity because it would only be conscious and be able to experience something after the activity resumed.

Of course, the room might be able to infer that what the person-in-the-room is doing results in its being conscious and, if the room is smart enough, then develop some confidence in this association (some have tried to formulate an objection to our argument using this point; we take this is up in the next section). Developing *some* confidence in the association is precisely what we claim consciousness researchers will one day be able to do, too – it will be one aspect of the science of consciousness. But this is a far cry from a satisfying explanation of how it is that consciousness supervenes.

Call consciousness's encapsulation property its *hermetical property*. One's consciousness is an omnipresent bubble in which we live our lives. We can't see anything causing it – we can't see our consciousness supervening on the NCC – because we can't step outside our consciousness and watch it come into existence via causation.

2.3.2 Handling an objection

2.3.2.1 *Types of phenomenal judgments*

When humans experience seeing something blue, they often form a belief that there is something blue, a belief of the form: "that's blue." This is one type of phenomenal judgment we can make. A second type occurs when we step back from our blue experience and note something about the experience itself, like "this is an experience of blue." A third type of phenomenal judgment happens when we again step back and pass judgment on our judgment about the experience. We might say something like "having sensations of blue is mysterious."

Now, consider what is really going on when we come to understand a supervenience relation. Let's look again at our moving car. Why does the car move? The car moves because of the movement of its parts: the pistons, the crankshaft, the drive train, and the back wheels. The movement of the car logically supervenes on the movement of the pistons, the crankshaft, the drive train, and the back wheels. (Remember, we are assuming the appropriate context.) Our reasoning is something like this:

1. seeing the car move (e.g., by seeing it drive by),
2. seeing the pistons, the crankshaft, the drive train move, and the back wheels move (e.g., by opening up the car's engine and transmission, etc.), and
3. making the supervenience inference and coming to understand how the causal relations of (2) could cause *per se* (1).

The supervenience inference in Step 3 is something of a puzzle. In general, within cognitive science, a detailed theoretical explanation of how this step is completed for a broad variety of cases is currently unavailable. But in the car case it is reasonably clear what happens: we make Step 3 by seeing how the up-and-down movement of the pistons gets translated into the longitudinal rotation of the drive train which in turn gets translated into the cross-wise rotation of the back axle, which, in turn, turns the back wheels.

What is needed to make the supervenience inference is to perceive the supervenience base, then the supervening behavior, and then inferentially put them together by inferring the supervenience relation. Inferentially putting them together requires connecting the relevant concepts (logically or otherwise – the car case is the logical version). Though this is not directly perceiving the supervenience relation, it is the next best thing and is sufficient for truly and satisfyingly understanding how the supervenience relation obtains.

Though the above outline of steps is crude, an objection denying the hermetical property of consciousness can now be mounted. Suppose that consciousness supervenes logically on some neural process. We can come to see, via inference, this supervenience relation, the objection goes, just as fully and just as robustly as we can in the case of the car. To see how, consider a specific case of consciousness, such as being conscious of a blue sky. Now, as closely as possible, let's follow the same steps as in the in car.

Step 1. At time 1, person P experiences blue.

Step 2. At time 2, P makes the second type of phenomenal judgment in which she is conscious of having an experience of blue. She experiences having an experience, saying something like "I was having an experience of blue at time 1".

Step 3. At time 3, P perceptually experiences (sees) NCC_{blue} (the supervenience base for her experience of the blue sky in Step 1).

Step 4. Now P has available for cogitating on the two relevant entities, the supervenience base, NCC_{blue}, and the supervening state, experiencing blue (which she got via the Step 2). She then draws the appropriate supervenience inference.

It follows, the objection continues, that there is no impediment to seeing via inference the supervenience relation. And though this is not directly seeing, it is enough to defeat our claim that consciousness has the hermetic property.[7]

2.3.2.2 *Why the objection doesn't work*

The objection only seems to work because it elides the very distinction we've drawn between seeing the supervenience relation in the car case and seeing it in the case of consciousness. The above four steps involving experiencing blue aren't anything like coming to understand the movement of a car logically supervening on the movement of its drive train. What is really needed to make the analogy with the moving car case, to make the objection work, is to see *how* NCC_{blue} could result in one's experience of blue. But this is precisely what can't be had because of the hermetic property of consciousness. So the objection begs the question against our view.

To come to see that the movement of the car supervenes on the movement of its parts, we trace the movement of the pistons to the movement of the drive train to the movement of the back axle. We decompose the movement of the whole into the movements of its parts. And that is just what a good reductive explanation is supposed to do. But we can't do that for consciousness. We can't decompose our consciousness into smaller bits, which we then map onto neural processes. If we break our conscious experiences down into bits, then we still have conscious experiences themselves. We can't get out of the bubble. The alleged supervenience inference in Step 4 above is really just a leap of faith. (This is another way of looking at Levine's 1983 explanatory gap.)

In the moving-car case we have a *vantage point*, at some remove from the car and its parts, from which to view the moving car and its moving parts. It is this vantage point that gives us the perspective that allows us to draw the supervenience inference. But in the case of consciousness, making a type two phenomenal judgment, being aware that one is having an experience of blue, is itself an experience, a bit of phenomenology. Hence, there is no appropriate vantage point from which to view the supervenience base and its supervening phenomenology. So, the objection fails.

2.4 An argument against naturalism

The Cartesian and zombie intuitions stand in the way of naturalism. To *explain* consciousness, the relevant concepts, those about our consciousness and those about our neural processes, must connect. And in order to do that, naturalism

must somehow defang the countervailing intuitions. The best way to defang these two intuitions is to render them persistent illusions. And the best way to do that is by deploying other perceptual information that has an epistemically superior etiology – one much better suited to science and scientific explanation. Unfortunately, the best such information – directly perceiving one's consciousness arise from its neural supervenience base – is logically impossible to get. This means that the best way to defang the Cartesian and zombie intuitions is unavailable.

Moreover, it is very unlikely that weaker methods of defanging will work. Any weaker argument will still have to be based on a cognitive-perceptual etiology more "impeccable" than our normal untutored cognitive apparatuses. That is, more than mere correlations between experience and some event in the world are going to be required to appreciate the supervenience base of consciousness. Otherwise, all we would have are intuitions, which are exactly what we are starting with and are exactly what is unacceptable here.

There are three ways to get something beyond correlation, something with the appropriate etiology. First is literally seeing the supervenience base as the supervenience base. This option is ruled out immediately with consciousness, as we have explained. Second is making a tight, circumscribed, small, and amply justified inference, as in the case of the car. This chapter rules out this option as well. We are left with the third option: embedding the correlations in a theoretical framework that permits and supports the second sort of inference.

However, in the case of consciousness, what we need is exactly some sort of embedding framework that allows reductive inferences. We can't assume that such a framework exists, or will exist someday, and then use that hypothetical framework to justify some sort of reductive claim, for the reductive claim would be part and parcel of the hypothetical framework. In other words, to choose this strategy would be to stand on our own tails. We would have to assume what we are looking for in order to claim that we have found it.

We conclude: if one has the zombie and/or Cartesian intuitions (and one likely has both, and almost certainly has the Cartesian intuition), one is probably stuck with them, and the explanations needed by and hoped for by naturalists cannot be got. Hence naturalism is untenable.

Chapter 3

The dismal prospects for naturalism

> I try never to think about consciousness. Or even to write about it.
>
> *Jerry Fodor*

Our argument against naturalism from the last chapter can be summarized as follows:

1. Naturalism will never produce satisfying explanations leading to understanding as long as the Cartesian and zombie intuitions are so easily adopted and held.
2. The best way to defang the Cartesian and zombie intuitions is by rendering them persistent illusions.
3. The best way to do that is by consciously perceiving one's own consciousness arise from its supervenience base.
4. That is impossible.
5. Weaker ways to defang the Cartesian and zombie intuitions are unlikely to work.
6. Therefore, it is unlikely that the two intuitions can be defanged.
7. Therefore, in all probability, naturalism is untenable.

Naturalistic readers are unlikely to be swayed by this argument. They will point out that the argument is probabilistic, depending, as it does in Step 5, on the surmise that weaker ways to defang the Cartesian and zombie intuitions won't work. If there were weaker ways to defang these intuitions, then the intuitions could perhaps be rendered impotent, or impotent enough, to clear the way for, e.g., materialism and hence an explanation of consciousness. In that case, naturalism would be vindicated.

Even if the probabilistic nature of Step 5 could be fixed, a naturalist might legitimately wonder whether all these appeals to "the best" produce merely one roadblock to naturalism, a roadblock circumventable by opting for weaker routes, which, though suboptimal, are nevertheless strong enough to sideline the countervailing intuitions enough to clear the way for naturalism.

Let us take up the naturalist's challenge. In Section 1, we reinforce premise 1. Then, armed with the conclusion that the Cartesian and zombie intuitions

really do block explaining how consciousness could arise from a physical (neural) supervenience base, it follows that either these intuitions have to be defanged directly or they have to be defanged some other way, say by finding a reductive, explanatory theory that finally reveals these two intuitions as illusions. These two options are exhaustive. We argued in Chapter 2 that the first option is impossible. This solidly shores up premises 2 through 4. That leaves the second option, which is basically premise 5. In Section 2, we show that there are no weaker ways to defang the offending intuitions.

3.1 The role of concepts and inference in supervenience explanations

Naturalism is held hostage by the Cartesian and zombie intuitions. These two intuitions make the needed conceptual connections between consciousness and the material realm impossible to fathom. The intuitions cause problems because of a special property of genuine explanations; viz., they require a conceptual connection between the explanans to the explananda. Hence, explaining consciousness in a satisfying way will require connecting "consciousness" with "neural processes" (or whatever). And, to be satisfying, this connection will have to involve a series of complex causal inferences. It is these inferences that will constitute our reductive understanding of consciousness. This understanding is grounded in conceptual connections that are aptly described as *logical* because they are aptly described as implication (of one sort). Any reductive accounting of consciousness that works will provide us with the necessary conceptual connections. That is, any satisfying explanation claiming consciousness is x will require that we conceive of consciousness as x, or that, given our other intellectual commitments, it makes sense that we conceive of consciousness as x (we will return to this point in Chapters 4 and 5).

A further point is that such inferences will have to reveal as *mechanistic* the causal path leading from the neural substrate to conscious experience (one can imagine that certain magical conceptual connections are postulated as being crucial, but these won't engender understanding because they aren't mechanistic, i.e., repeatable given a well-defined set of precipitating conditions). Anything less than all of this and it will be mysterious as to why and how consciousness appears in the brain when it does and with the content that it does.

All of our satisfying, reductive theories work this way: they are satisfying only if the reduced and reducing phenomena are logically related (consider again, our example of reductively explaining a moving car from the last chapter). Another way to put this is to say that the connection between reduced and reducing phenomena has to be *epistemically compelling*. Such compellingness has to be rendered as logical inference. Thermodynamics is a well-known and good scientific example of the logical nature of reductive explanation. A modern and particularly interesting example is the logical relation between the virtual machines making up the software one uses and the very low-level hardware on which it supervenes. Here is an extended example.

Biology depends on Darwin's theory of natural selection as one mechanism for evolution. Suppose biologists knew about the existence of genes and postulated that the transmission of certain genes and not others was the underlying way natural selection worked. But suppose that that is *all* we knew and all we were even marginally confident of. Suppose that we were completely in the dark about what genes actually do relative to the organisms they code for, i.e., we didn't know that genes are strands of DNA, that they code for proteins, and it is these proteins that are the building blocks for the bodies and behaviors of organisms. Suppose, also, that we knew nothing of the distinction between phenotype and genotype. In short, suppose we didn't know how bodies and behavior arose from and depended on the microscopic workings of genes, proteins, and associated molecules – we were completely ignorant about developmental, cell, and molecular biology.

In such a case, we would have two disparate theories: the theory of natural selection (more fit organisms produce more offspring, and fitness is relative to an environmental niche), and a weak, vague, proto-theory of genes (genes are probably complex molecules involved somehow in evolution). In such a case, we could not see any compelling connection between the workings of genes and the formation of bodies. We might surmise that, e.g., wooly mammoths were woolly because their climate was getting colder (paleo-geology would be needed here), and that a mammoth with more wooly hair would *ipso facto* be healthier and hence able to have more offspring. But we wouldn't know how genes code for wooliness, nor how gene variation codes for differences in wooliness. We would just figure that genes were implicated somehow.

In this case, natural selection's connection to the molecular world would essentially be mysterious to us. We would believe that genes were involved in evolution for purely abductive reasons. But we would not understand in any satisfying sense how genes and natural selection worked together to produce change in organisms. Our understanding of evolution would be very thin.

This situation is historical. It is the exact situation that occurred after the rediscovery of Mendel's genetic experiments around the beginning of the twentieth century. The connection between genes and natural selection was so

poorly understood that there were *two* theories of evolution: one based on Darwin's notion of natural selection, and one based on Mendel's notion of genes. The two camps were called "Darwinists" and "Mendelians," and, in an ironic twist of cruel fate, Darwinism was losing: most biologists believed that genes were the key to evolution and that natural selection was a very minor force, if even that, in evolution (one might see in this bit of history scientists' general penchant for reduction). Seeing no clear causal connection between the workings of genes and the shapes of bodies, the function of body parts, and the roles of behaviors, early twentieth century biologists thought that there were two theories, and only one of them could be correct (or only one of them would be central), and that the one that was correct probably wasn't natural selection.

Of course, the situation has changed dramatically since then, beginning, in the 1920s and 1930s, with the work of R. A. Fisher, J. B. S. Haldane, and Sewell Wright. We now know that both the principles of genetics and natural selection are needed for understanding evolution. We are well on our way to understanding how it is that DNA actually codes for various proteins and how these proteins interact with the environment and the DNA itself to form bodies. At the same time, we are coming to appreciate that our original conceptual of genes as biological atoms was mistaken and that genes qua bits of DNA that code for specific phenotypic traits probably exist only rarely. More common are stretches of DNA that interact with each other and the local environment. As we learn more and more about the causal interactions among DNA, RNA, the local cellular environment, and so forth at the micro-level, we are also learning new ways to conceptualize genes and evolution at the macro-level. These new conceptualizations then reflect back on how we frame the underlying causal interactions.

Though still incomplete, our current biological theorizing is nevertheless rich, profound, and satisfying. And at its most basic level, what makes it satisfying is that it provides us with a robust and compelling mechanistic and conceptual connection between the workings of DNA and the shapes and functions of bodies.

By analogy, *if* we were to develop a satisfying explanation for consciousness, then it too would draw on the details of causal connections to frame the conceptual connections we would make between consciousness and its supervenience base in our brain. But that "if" is insurmountable. Regardless of how any attempted explanation of consciousness might go, the Cartesian and zombie intuitions drive a wedge right between the two kinds of concepts that need connecting. And that wedge is here to stay as long as these two intuitions are *so* intuitive, *so* easily believed. These two intuitions make the needed connections between consciousness and the material realm impossible to fathom.[1]

3.2 The conceptual impasse

No mere amassing of correlational evidence is ever going to result in an explanation of how consciousness arises from its neural substrate; it will merely establish that it does, in some detail, hopefully. Defanging the intuitions that blocks our inference from correlation to identity *requires* satisfactorily explaining consciousness. But explaining consciousness in a satisfying way requires that the concepts of consciousness and its substrate be tied together in fundamental ways.

We have an argument that no conceptual connection between the relevant concepts can be had. This argument requires some setting-up. We need to introduce the technical notions of subjective and objective concepts, points of view, and some logical machinery governing the inferential relations of these notions.

For us, *concepts* are token mental representations active in working memory. We will also consider token but nonactive representations in long-term memory to be concepts. But, when we use the word "concepts" alone, we will always mean active concepts in working memory. When discussing long-term memory concepts we will always flag this fact explicitly. Working memory concepts are constituents of beliefs and other propositional attitudes (which are also tokens). Concepts are involved in many kinds of mental processing, and support crucial epistemic capacities, notably categorization and recognition.[2]

Concepts are decomposable and their parts have structure. The nature of conceptual parts and their structure is the subject of a large, ongoing research project to find a robust theory of concepts that fits all the perplexingly varied data (see Murphy 2002, for an introduction). We intend to steer clear of this controversial topic as much as possible, so we will remain neutral about what these parts are and how they are put together. But we do need to distinguish between types of parts: those parts involved in the semantic decomposition of concepts and a component related to an aspect of consciousness (this distinction will perhaps prevent us from avoiding controversy). We will explain this distinction shortly.

We define *subjective* and *objective concepts* as concepts "participating" in or "associated with" subjective and objective *points of view*. In this, we are following Nagel, who pointed out that it is beliefs and other attitudes that are the

primary carriers of subjective and objective points of view (1986: 4). Points of view are properties or aspects of one's occurrent consciousness (which seems to imply that points of view are somehow associated with working memory). We assume that at least some concepts in working memory are part of our occurrent consciousness and that no consciously entertained concept is point-of-view free. (If one considers, as we do, some long-term memory representations to be concepts, then those concepts are point-of-viewless – but, crucially, they are also inactive). Going the other way, there is also no such thing as a conceptless point of view. One cannot just have a brute, bare point of view. All points of view involve activating some concepts.

It would be best now, to define *points of view*. Unfortunately, rigor proves elusive here. That said, we don't want to leave an understanding of points of view completely to unexercised intuition. Just as ostension works to locate or pick out consciousness as a topic of study across different individuals, so too will ostension help to pick out what we mean by points of view. Points of view, even objective points of view, really do involve *views*. It's not just a metaphor or way of talking. True, the views referred to are made with the "minds eye," not with any actual perceptual organs. And, true, this latter phrase seems to suggest that the phrase "points of view" in fact is just a metaphor. But the problem is with the "just". They aren't *just* metaphors. Points of view really do somehow involve something quite like perceptual consciousness, giving us different views on our world.[3]

Nagel, in his seminal studies of objective and subjective points of view (1979, 1986), suggested that pure subjective and objective points of view are really endpoints on a (noncontinuous) spectrum. For Nagel, points of view are actually more or less objective or subjective relative to one another (see, his 1979: 206). Our view is different. We think that points of view are strictly binary; they are like the set {0, 1}, with no in-between gradations. What Nagel described as ever more objective points of view are better described as ever *wider* objective points of view. Such points of view are *not* more objective, they simply encompass a wider perspective that is already as objective as it can be. (We suspect, but don't know, that the situation is asymmetric: subjective points of view don't seem to be increasingly narrow.)

Our notion of how points of view work contrasts with Nagel's with respect to the way points of view change. To understand Nagel's idea, it is best to begin with one's most subjective point of view. Suppose you are having a square-shaped, red phenomenal experience. The subjective point of view must leave unspecified how it is you are having this experience. It is purely your experience. Any specification of something causing your experience would be

describing something accessible to others and hence something objective. All of the particulars that go into making this red square experience *your* experience – your specific sensations and associations, etc. – are what makes this experience of yours subjective. Whether there is something in the world causing it, whether there is a world at all, is irrelevant – you are having a pure subjective experience, period. And your point of view in that experience is the subjective point of view.

A point of view, according to Nagel, becomes more objective to the extent that it doesn't rely on your specifics: the more accessible to a variety of conscious agents the information content of a point of view is, the more objective it is. If your square-shaped, red phenomenal experience is caused by, say, looking at a large, red square on a wall, then that point of view is more objective since the red square is an object external to you and hence is more public and accessible to others. Still more objective points of view can be got, according to Nagel, by considering the red square not as causing red sensations but as reflecting light of wavelengths between 630 and 750 nanometers, since more kinds of conscious creatures have access to these measurements than to the red experience (colorblind people for example). And, since nanometers are a human form of measurement, a still more objective point of view of this situation can be obtained by considering the square as reflecting light of some wavelength specified by whatever measurement a perceiver of the square might happen to use (see Nagel 1979: 206–207, 1986: 4–6).

Nagel considers the switch from subjectivity to objectivity to be accomplished by standing back and including one's own subjective view within a new view. For him, this is definitional. Furthermore, he claims that this standing back can be done over and over again, as one relinquishes more and more details that make the point of view even slightly personal. On Nagel's view, the nature of the subjective and objective is one of concentric circles where the most subjective point of view is at the center and each more objective point of view includes the next innermost subjective "circle" as a proper part.

However, we disagree with Nagel's analysis. The three objective points of view – the red square on the wall view, the nanometer view, and the generalized measurement view – are not each more objective. They are simply each wider in scope. The subjective point of view comprises the subject's unique, one-time only, first-person perspective on the world. (It might be thought that the one-time only aspect could be relaxed, but it can't. Reidentifying the same experience across distinct times is a kind of objectification. It is the minimal width objectification, since the isolated individual need be the only one involved. But it is a kind of objectification nevertheless.) Objectivity is ob-

tained immediately and completely by a sort of *reification*, an *objectification*. This reification posits an external *cause* that is responsible for the subjective experience (we use the word "external" in this context to mean external to the subjective experience, not necessarily external to the person having the experience – the cause of a leg pain could be in one's leg, for example). This cause is some external object or process that is potentially communal and public, even if the "public" is the individual at a different time. (Note, our use of the terms "communal" and "public" are slightly nonstandard. We mean them to include the standard definitions, but we also include individuals at different times picking out the same or very nearly the same experience.) Subjective states are never communal and public, not even potentially. But their objective counterparts are.

Now, it should be obvious that this shift from an in-principle-not-public, subjective point of view to a potentially public, objective one can only made once. From here, wider perspectives can be achieved, but they are not more objective because they are not more public; the objects responsible for subjective experience cannot get more public or more external. It is true that more and more *of* the public, of the community, might observe the objectified cause of the subjective experience (the community can eventually expand to be the universe), but this does not make the objective point of view of that cause more public, more external; it is just more widely observed. (We note here that objective viewpoints of greatly differing widths frequently are viewpoints of different things that are causally related in the external world. The three objective points of view – the red square on the wall view, the nanometer view, and the generalized measurement view – exhibit this property.)

So, objectivity is essentially communal or potentially communal, even if the community is a single, diachronic individual, or is imagined, virtual, or abstract. One's viewpoints can be widened by including larger and larger communities, but that doesn't make the viewpoints *more* communal. It merely makes the communal view wider. Hence, the wall view, the nanometer view, and the generalized measurement view are each wider, but equally objective points of view.

Note that objective points of view allow us see our subjective experiences as caused by events in an external world. It is by adopting an objective point of view of one's square-shaped red experience that one can say things like "That red square on the wall is causing my sensation of red." So when we adopt an objective point of view, we are able to include subjective aspects of ourselves (or perhaps subjective versions of ourselves) as part of the community.

The subjective point of view is also called the *first-person perspective*, and objective point of view, the *third-person perspective*. Our conclusion can be also gotten by considering these perspectives. There are no degrees between first and third-person views. Nothing is a little bit first-person but mostly third-person, for example. Intuitions to the contrary are in fact based on noticing that one can quickly switch between the two points of view (or perspectives), and even switch between different widths of objective points of view and the subjective point of view, keeping them all easily in working memory. Nevertheless, first and third person viewpoints are strictly binary.[4]

This is a good place to note that it seems likely that one can entertain or hold only one point of view at a time. Again, one can switch quickly between them, but they are held or entertained serially.

It is important to be careful here. One's objective view of the world is not less conscious; it does not involve less of one's consciousness. It is still *one's* view. It is just that this view *objectifies* one's first-person phenomenal experiences by construing them as caused by something in an external world.[5]

Being able to adopt wider and wider objective points of view is crucial to living in a community, to having communication, and to doing science, etc. Each wider view includes the previous, less wide one as a proper part. This continual widening leads to a kind of hierarchy of points of view, with the unique, subjective point of view in the center. Such widening perspectives can proceed well beyond being human. As our objective point of view widens, we come to a view of our world that is neither from any special place within that world, nor from any special place at all. Finally, at its widest, one's point of view is, in a sense centerless, from nowhere (of course, we humans probably never achieve such a wide objectivity, but, rather, approach it asymptotically). As Nagel suggested, this widest point of view is, in an important sense, a *view from nowhere*.

Though we disagree with Nagel that points of view are relatively objective, we do adopt a related claim: an objective point of view is always relativized to a certain, specific subjective point of view; a given objective point of view is objective only relative to a specific subjective point of view. Furthermore, wider objective points of view are relative to their narrower cousins.

We can now present our (very) rough definitions of subjective and objective points of view. A point of view is *subjective* if it crucially involves only phenomenal properties unique to a subject. A point of view is *objective*, relative to that subjective point of view, if it posits external objects and processes, at least in principle accessible to others, which are causes of the associated subjective point of view.[6]

(We will not discuss in this chapter, the common but difficult case of adopting another's point of view, though we will touch on one aspect of this topic in the next chapter. Adopting another's point of view is an attempt at adopting another's subjective point of view, not adopting a more objective point of view. Adopting another's point of view works best when humans adopt the points of view of other conscious agents, as when we adopt the points of view of each other or our pets or other mammals. But, it is not clear we can adopt the points of view of phylogeneticially distant animals. These cases are difficult because, as pointed out by Nagel (1974), the more different the entity is from us the less it is clear that we are adopting another's point rather than imagining ourselves being in the role or position that that other thing is. Can we adopt a bat's point of view, or can we merely imagine what it might be like for us to hang upside down and hunt for food by flying around and listening for the high-speed reflected clicks that indicate a bug's presence? These two seem very different.)

Now we can more deeply relate points of view to concepts.[7] Imagine again that you are having a square-shaped, red experience. When you *attend* to this experience and form the belief that you are having a square-shaped, red experience, you have active in your working memory the concepts for squareness and redness (along with others).[8] These are *phenomenal concepts*, concepts that refer to the relevant parts of your phenomenal experience.[9] Your subjective point of view includes concepts that are utterly idiosyncratic to you. We can think of such concepts obtaining whether or not there is a world out there causing your red square experience, or alternatively, that for this kind of phenomenal concept, that there is an external world causing your experience is irrelevant to the concepts since they refer to sensations and phenomenal properties. We call such phenomenal concepts *subjective concepts*.

Your objective point of view (at some specified width), on the other hand, includes different phenomenal concepts that are publicly communicable or shareable (via language, usually). Such concepts do not rely exclusively on your specifics. Your wall view, nanometer view, and generalized measurement view each contain concepts that refer in certain ways to various objects and properties in the external world. It is these objects, conceived of with wider and wider perspectives, perspectives that are less and less tied to human perceptual organs, which make the concepts that pick them out objective. Such phenomenal concepts are *objective concepts*.[10]

Given our analysis of subjective and objective concepts it is obvious that no subjective concept ever logically implies an objective concept. Being appeared to red squarely never implies that there is a red square out in the world causing the appearance because it is possible that the appearance is not caused by a red square (one could be hallucinating or dreaming, for example). From this it follows that subjective and objective concepts have different content simply because the second type of concept but not the first implies the existence of objects in the world; objective concepts imply a world, whereas subjective concepts imply experiences. Some might find it next to impossible to see subjective concepts as objectively inert, but they are, nevertheless. When the world is well-behaved, then subjective and objective concepts refer to different aspects of the same thing (though note, subjective and objective concepts do not have the same referent). So, subjective concepts derived by attending to red square experiences typically refer to redness and squareness – experiences usually caused by red squares existing in the external world. But this is merely a happy, very reliable and so frequently relied on, state of affairs. There is no logical necessity about it. It is an inductive leap we make because we live in a well-behaved world. That we find the inductive step irresistible doesn't diminish its inductiveness.

Going the other way is a bit trickier. If you have the objective concept *red square on the wall* or *that red square on the wall* then that does entail for you that you have the subjective concept *red square*, for as we mentioned above, objective points of view are relativized to subjective ones (to be precise, one has to assume a realism here – the sort of realism that posits a mind-independent world that is very similar to our phenomenal and mental world; perhaps the reader will grant this). But notice that this inference only works for each separate conscious individual. You and Smith may both have an objective concept of a red square on the wall, and you each may then correctly infer that you are having a red square experience with the appropriate subjective concepts, but you can never logically infer that Smith has the appropriate subjective concepts (nor he you) since you cannot be sure that Smith is having any experiences at all: Smith might be a zombie. So inferences from objective concepts to subjective one's go through for you, but do not go through for others relative to you. This is just another way of saying that leaving the objective point of view and going to the subjective one is something you can only do for yourself. You can never do this for anyone else.

It is important to point out that this logical problem for subjective and objective concepts is only a problem for the science of consciousness. When we reduce, say, temperature to mean molecular motion, we have the same, objective, point of view on both (indeed, we usually have the same objective width, too). Hence the point of view is invariant between the reducing process and the reduced process, hence it can be, and is, ignored.

Now, with all of this in place, we can easily make our argument that the nature of subjective and objective concepts prevents ever defanging the Cartesian and zombie intuitions, prevents ever closing the explanatory gap.

Since directly seeing consciousness arise from its supervenience base is impossible, defanging the intuitions is left as a scientific enterprise. This requires seeing how phenomenal states or experiences logically supervene on their physical supervenience bases for all kinds of conscious entities. Seeing how phenomenal states logically supervene requires that there be a logical relation – logical entailment – between the two relevant kinds of concepts: subjective concepts and objective concepts. But there is no such logical relation, except within the case of a specific conscious entity where certain objective concepts do imply certain subjective ones – but this is useless for the science of consciousness, which like all sciences has to be communal. So defanging the intuitions is not in the cards.

It is interesting to see how really bad off the situation is from a logical point of view. Let Bartleby be a conscious agent. Suppose that Bartleby is looking at a red square on a wall. Call Bartleby's red square experience, *R*. Consider a subjective concept, *S*, formed when Bartleby attends to *R*. And consider the corresponding objective concept *O* in Bartleby's mind. *S* and *O* refer to different things: *S* to Bartleby's *R*, and *O* to the red of the square on the wall. Call this external red, R^*. *R* is something in Bartleby's mind, whereas R^* is something in the world; R^* is public, but *R* is not. As a heuristic, we might pick out R^* using the English language description "light of wavelengths between 630 and 750 nanometers". (We are not completely sure how to characterize R^*. It seems as if it must be related to *R* somehow. Perhaps it has *R* as a major constituent (or vice versa). Perhaps *O* uses *R* essentially, but construes it from an objective point of view. We don't want to get bogged down on this issue, however. All that matters for our purposes is the point that *S* and *O* refer to different things.) These two concepts, *S* and *O*, would participate in different beliefs. *S* in beliefs like "That is red [where the demonstrative picks out Bartleby's phenomenal state, *R*] or "I am experiencing red", and *O* in beliefs like "That square on the wall there is red". Now let Jones be a neuroscientist observing Bartleby by using some high-tech neuroimaging machine. Suppose that Jones forms the objective concept, *B*, which refers to Bartleby's brain state – the one Bartleby is in when she has *R*. There are no logical entailments anywhere in sight that are of any use to the science of consciousness. *S* doesn't imply *O* even for Bartleby and *O* implies *S* only for Bartleby. This is true because, as we pointed out above, *S* and *O* are so different, and they are so different because they each participate in or are associated with different points of view. So just considering Bartleby,

there is no closing the explanatory gap: her subjective concepts and objective concepts inhabit logically different realms. *S* and *O* are related, of course, as we pointed out above: *S* refers to *R* and *O* refers to R^*. But these two are only related relative to Bartleby. Consequently, even for Bartleby, the relation between *R* and *O* is mysterious, at least on the logical level.

For Jones, the situation is even worse. *B* and *R* are not only from different points of view, but, *B* doesn't even have the right referential semantic content: *B* refers to brain states. So, unlike *S* and *O*, which refer to related phenomenal properties, *R* and R^*, but differ in point of view, *B* shares nothing with *S* and *O*. So the logical relation between *B* and *R* is even further apart.

It might be objected here that to be explanatory the science of consciousness need not relate subjective and objective concepts, that the science need only relate differing objective concepts, say the objective concepts of someone perceiving a red square and some neural processes on which that perceiving supervenes. But this is incorrect. To be explanatory in a satisfactory way *is* to relate one's subjective concepts to one's objective concepts. One has to inferentially relate, e.g., one's red square experience to one's neuro-processing concepts on which those experiences supervene, otherwise, there will be no reason whatsoever to regard the propose explanation as anything more than just a correlation. If neuroscientist Jones relates Bartleby's neuroprocesses with, e.g., Bartleby's reports of experiencing a red square (or any other behavior of Bartleby's), the best such relating can aspire to is correlation. And correlations, as we have already noted, are not explanations. This is a good place to allude to the next part of the book.

It is clear that explanations of consciousness are not in the cards because the relevant concepts cannot be put in their proper relation for explanation. But it doesn't follow from this that the science of conscious is stymied. It merely means that the science of consciousness won't be explanatory. There will be more on this in the next two chapters.

So, finally, back to defanging the Cartesian and zombie intuitions. Without directly perceiving consciousness supervene (which you cannot do), *no* weaker way of defanging these intuitions will suffice, since all of those ways will require some method of logically hooking up the relevant subjective and objective concepts so that reductive inferences can be made, and, as we just shown, that is impossible. So the Cartesian and zombie intuitions are here to stay, as is the explanatory gap.

Of course, our argument that subjective and objective concepts are logically unconnectable allows for a strong, more direct conclusion. As we pointed out in our defense of premise 1, above, a deep, satisfying understanding of con-

sciousness will be a reductive understanding. And a reductive understanding will require intuitively compelling logical inferences from the subjective to the objective (and vice versa), which we've just shown cannot be got (except in an individual, in the case of going from objective concepts to subjective ones). So a reductive understanding of consciousness will always elude us. So naturalism is untenable.

We finish the defense of our main argument by pointing out that steps 6 and 7 are merely unzipping the modus tollenses. So we have our conclusion: Naturalism is untenable.

3.3 The improved argument

Here now, is our improved argument in final form:

1. Naturalism will never produce satisfying explanations leading to understanding as long as the Cartesian and Zombie intuitions are so easily adopted and held.
2. The only way to defang the Cartesian and zombie intuitions is by rendering them persistent illusions.
3. But the intuitions are here to stay.
4. Therefore, naturalism is untenable.

Here is a short version of the argument spanning the last two chapters:

1. You cannot witness consciousness supervene on its supervenience base.
2. You cannot draw any supervenience inference that consciousness supervenes on its supervenience base.
3. But either 1 or 2 has to be possible if naturalism is to be tenable (since naturalism requires consciousness to supervene on something).
4. So naturalism is hopeless.

If one lets "supervenience base" be the physical realm, then this argument refutes scientific materialism – the view that a material explanation of consciousness is in the offing. If, following Chalmers, one lets "supervenience base" be some special, protophenomenal property of matter, then this argument shows that scientific dualism is not in the cards.[11] These are the two contenders.[12] Hence, a scientifically satisfying explanation of consciousness is not in the cards.

PART II

Aspects of a science of consciousness

Chapter 4

How to avoid being a mysterian

> . . . the problem of consciousness no longer seems intractable.
>
> *Stanislas Dehaene and Lionel Naccache*

4.1 The lure of the mysterian view

In one of the most widely cited articles in the philosophy of mind, Thomas Nagel tells us that we can't know what it is like to be a bat (Nagel 1974):

> Even without the benefit of philosophical reflection, anyone who has spent some time in an enclosed space with an excited bat knows what it is to encounter a fundamentally *alien* form of life.

We can't truly know, according to Nagel, what it is like inside someone or something else's head. We can guess, and quite often guess well (apparently), but we are also often wrong to varying degrees. The problem, of course, is that we don't have any way of provably accurately and fully translating the external cues we get into what someone else's internal states actually are.

Nagel's is not solely a claim about the underdetermination of theory by evidence, though it is that. It is a claim about the limits of our imaginations. We can't know what it is like to be a bat because we can't imagine very well the dark, sonar world of bats. Even if we can imagine what the bat clicks (the echoes) sound like, we find it virtually impossible to imagine perceiving our environment in any detail using echoed clicks, let alone imagining what it would be like to fly around rapidly in that environment. Our species-specific capacities determine what we can and do experience.

The limits of our imaginations are somewhat fuzzy. It is well known that it tends to be easiest to imagine what it is like to be something else much like us. In our own case, it is easiest for us to imagine what it is like to be early twenty-first century, Anglo-American philosophers of mind living in the eastern United States. It really helps if the other philosophers are down the hall. In general, it is easiest for humans to imagine what it is like to be other humans.

From here, humans' capacity to imagine what it is like to be something else degrades: it is somewhat easy to imagine what it is like being a dog, somewhat harder to imagine what it is like being a horse, and hard to imagine what it is like being a lobster, assuming it is like anything. What is curious about our human capacity for imagining is that when any pressure is put on it, i.e., when detail is required, it becomes hard to imagine what it is like to be even a human very similar to ourselves. So, for example, if much detail is required, Dietrich finds it hard to imagine what it is like to be Hardcastle, and vice versa. The more detail one asks for, the more difficult it becomes to imagine what it is like to be something or someone else. What makes bats and their ilk interesting is that, without asking for much detail at all, it is still very difficult to imagine what it is like to be one.

It seems clear that the gradation of successful imagining – from somewhat successfully imagining what it is like to be another twenty-first century, Anglo-American philosopher of mind living in the eastern United States to failing to imagine what it is like to be a bat – itself indicates that we don't really ever imagine what it is like to *be* another. Rather, what we do is project our own knowledge of our own experiences onto the other. This really only works if the other is like us.

Of course, language plays a large role in helping us imagine what it is like to be another. If the other can describe for us what is it is like to be them, we can go some ways toward imagining what it is like to be them. But even here, language is only going to provide a shadow of what the other is really experiencing. Nagel asks us to consider what it is like to a blind and deaf person, or vice versa (1974). It is clear that in this case, language is not going to help much. Hardcastle can describe to Dietrich what is like to give birth, but this helps Dietrich only dimly imagine what it is like.

Nagel's conclusion from these observations is strong:

> Reflection on what it is like to be a bat seems to lead us, therefore, to the conclusion that there are facts that do not consist in the truth of propositions expressible in a human language. We can be compelled to recognize the existence of such facts without being able to state or comprehend them (1974).

There is a way to turn this difficulty in imagining what it is like to be another into a deep skepticism about a science of consciousness. The argument is this:

Argument 1

1. That we can't consistently well-imagine another's subjective experiences means that in general, the subjective, first-person experiences of others are inaccessible to us.
2. This inaccessibility means that first-person experiences are not at all translatable (or only very weakly translatable) into objective, third-person descriptions (if the translations were robust and easily obtained, others' subjective experiences would be accessible to us).
3. But science requires robust, objective, third-person descriptions, and specifically, a science of consciousness requires robust objective, third-person descriptions of all manner of experiences from all kinds of conscious entities because it these experiences that form the explananda, the raw material that the science will explain.
4. Therefore, there won't be a science of consciousness.

Argument 1 proceeds from inaccessibility. Frank Jackson (1982) uses knowledge. He presents us with a thought experiment from which we can draw a similar science-obliterating conclusion, except instead of asking the reader to note how hard it is to imagine being a bat, Jackson asks the reader to imagine being Mary,

> a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like "red" and "blue", and so on.

Then Jackson asks us to imagine:

> [what] will happen when Mary is released from her black and white room or is given a color television monitor. Will she learn anything or not?

It seems she will. She will learn what the color red is, for example, upon first seeing ripe tomatoes in the flesh. Also, her subjective concept of red (and no doubt color) will change dramatically. Consequently, knowing all the physical facts that there are is pretty obviously not sufficient for knowing about red or about color. The explanatory gap again rears its ugly head: we simply cannot know how physical facts give rise to phenomenal facts. And again, since to truly and satisfyingly explain consciousness, we will need to bridge the gap, it follows that an explanation of consciousness is not in the cards. Since explaining consciousness is what a science of consciousness is all about, we see that a science of consciousness is not in the cards.

When one notices that the epistemic gap in both Nagel's and Jackson's arguments is due to the accessibility/inaccessibility of other's conscious states, it becomes clear that the two are related.[1] So, in general form, the mysterian argument is this:

Argument 2

1. All scientific theories, by their very nature, describe the world from a third-person, objective perspective; this is true of both explanans and explananda.
2. Conscious states, by their very nature, are experienced from a first-person, subjective perspective; as explananda, they are inaccessible.
3. Therefore, scientific theories of consciousness can never capture all the relevant (or even all the important) aspects of phenomenal experience.
4. Therefore, no scientific theory of consciousness is possible.

Or, the mysterian argument can be summed up simply:

Argument 3

1. The Explanatory Gap (the one concerning consciousness) is ineluctable.
2. But science is in the business of closing explanatory gaps of each and every kind.
3. Therefore, there will be no science of consciousness.

For some, the lure of mysterianism does derive from saving from the onslaught of science what is central to human existence: our consciousness. But this needn't be the lure. For many, and arguably, for both Nagel and Jackson, the lure of mysterianism is simple honesty. The consciousness of another human or animal *is* inaccessible to a third party; there really is an explanatory gap. It is quite intuitive that this fact should have negative scientific repercussions.[2]

4.2 Problems with mysterian arguments

On the other hand, it seems preposterous that something as important as consciousness should be beyond the reach of science. This faith is bolstered by fatal problems with Nagel's and Jackson's arguments, and with the mysterian arguments we just rehearsed.[3]

First, let's consider Nagel's and Jackson's arguments independently of their mysterian force. Nagel's argument requires the reader, along with Nagel, to

fail to imagine something: the phenomenology of being a bat. But Jackson's argument is just the opposite of Nagel's: he requires his reader to *succeed* in imagining what it is like to be Mary, and then to imagine what it is like to be Mary stepping out in the sunlight for the first time ever. Note how uncomfortably these two arguments comport with each other. Nagel's argument requires a failure of imagination. Nagel guarantees this by using something different from us – a bat. Jackson's argument, in contrast, requires the success of imagination. Jackson seemingly guarantees this by using a human. But consistency seems to require Nagel to object to Jackson's argument on the grounds that Mary is actually radically different from us, so different that Jackson can't reasonably expect us to imagine what it is like to be Mary at all. Mary is not just a brilliant neuroscientist (hard enough to imagine) but she knows all the neurophysiological facts there are to know about vision; facts we don't even approach knowing. This is next to impossible to imagine; it is not even clear it is meaningful. It is also quite difficult to imagine living in a black and white room or environment. So, imagining being Mary is at least as difficult imagining being a bat. For Nagelian reasons, Jackson's argument fails, it seems.

Consistency requires the reverse from Jackson: he has to object to Nagel that he (Jackson) can perfectly well imagine what it is like to be a bat, since after all, he (Jackson) can imagine what it is like to be Mary, who, in truth, is really quite strange. So, for Jacksonian reasons, Nagel's argument fails, too.

The two arguments mutually self-destruct.

But what about the mysterian arguments we developed from Nagel's and Jackson's argument. The main problem with Argument 1 above is premise 3. We grant that a bat's consciousness is inaccessible to us. We also grant that imagining what it is like to be a bat is difficult. But this doesn't imply that imagining another's consciousness is scientifically useless, for all we have to imagine is *that the other is conscious*. Once we do this, we can let our science provide us with the necessary correlations governing the details of the other's consciousness; we needn't imagine these details with any success at all. In short, the science of consciousness does *not* require robust objective, third-person descriptions of all manner of experiences from all kinds of conscious entities; it can get by just fine with thinner, vaguer, third-person extrapolations of these experiences.

Consider again the quote from Nagel we used above, only this time note our editorial addition:

> Even without the benefit of philosophical reflection, anyone who has spent some time in an enclosed space with an excited bat knows what it is *like* to encounter a fundamentally *alien* form of life.

In order for Nagel's argument to get off the ground, he has to get his reader's intuitions going his way. He accomplishes this by asking his reader to imagine something he (Nagel) as trouble imagining. Nagel imagines that his reader will have the same trouble he has in imagining what it is like to be a bat. Nagel assumes that he knows what it is like to be us having trouble imagining what it is like to be a bat. But the details of this imagining are irrelevant to his argument. All Nagel needs to get his argument going is for his readers to have trouble imagining what it is like to be a bat. Nagel can be quite confident that we will have such trouble without having to imagine in any deep detail what it is like to be us failing to imagine what it is like to be a bat. In fact, perhaps Nagel can't really imagine what it is like to us, Dietrich and Hardcastle, failing to imagine what it is like to be a bat, for bats are probably stranger to Nagel (an inhabitant of New York City) than they are to us, who see bats all summer long and put up bat houses from time to time in forests. So his argument can succeed merely on the assumption that there is something it is like to be us failing to imagine what it is like to be a bat; the argument doesn't require knowing what that something is. *Just so* with the science of consciousness: all it needs is the assumption there is something it is like to be the conscious entity under investigation, or conscious entities, in general; it needn't know or explain or have access to any details about what it is actually like to be that entity. The existence claim is enough.

A robust science of consciousness will have most of its origins in studying humans, and not just for practical reasons like they can report what they experience, but also because humans are doing the research, so a deep similarity is assured. Once this is accomplished, there is every reason to believe that the scientific findings could be extended to other species.

But will this science tell what it is like to be a bat? That 's the question many think the science has to answer. We have just argued that the science doesn't need to answer that question.

Turning now to Jackson: there is a logical infelicity in his argument. As he runs his argument, Mary only concludes that her neurophysiological knowledge isn't sufficient to know what color red is after she is allowed to leave her black and white room. So described, it is not just hard to imagine Mary or imagine being Mary, it is impossible. Jackson asks us to imagine living in a black and white room knowing all there is to know about the neurophysiology of vision without ever concluding that something was wrong, without ever concluding that this knowledge was woefully inadequate to explain consciousness, visual consciousness, in particular. Black and white are colors, too. Mary had to know that her knowledge of the neurophysiology of vision was completely inadequate to explain black and white vision. In fact, as long as we're at it, let's assume that Mary knows all there is to know about *all* human neurophysiological functioning, from sensorimotor feedback loops involving sensory and motor neurons to the workings of the spinal cord, brain, etc. etc. With this assumption, it is clear that, even in her black and white room, Mary will know that something is sorely amiss. All her knowledge doesn't imply *a priori* anything about any aspect of her consciousness – about how apples taste to her (which of course look gray to her), about the feel of her keyboard, about the sound of her voice, etc. She doesn't have to leave her room to conclude that neurophysiological knowledge is hopelessly inadequate to the task of explaining consciousness. She can conclude that consciousness is mysterious without ever leaving the comfort of her black and white environment.

From here, it is obvious that we can jettison the black and white room part of the argument and the transition to a colored environment as showmanship. Mary can be a neuroscientist, living in a standard, colored environment, still knowing all there is to know about neuroscience, and still conclude that neurophysiological knowledge is useless to explaining her consciousness. Jackson's argument goes through without all that black-and-white-to-color razzle-dazzle. So why does Jackson go to all the trouble of setting up his argument in the elaborate way that he does? For effect. This problem highlights the intuition pump nature of his argument.

Don't misunderstand us. We are not objecting to Jackson's argument because it is an intuition pump. Intuition pumps have their place. We use them ourselves. We are making a point about imagination and science. Human imagination is plastic and malleable. With the appropriate exercise and the appropriate guide, it can go virtually anywhere, conjure virtually any notion or scenario. But imagination can also get into ruts, and it can also be stymied. This is especially true where consciousness is concerned. Having lived in conscious bodies all our lives, we humans tend not to notice how genuinely puzzling it is. Jackson needs to deploy an intuition pump in order to get researchers to see that there is more to understanding consciousness than understanding the neural facts it supervenes on (perhaps only naturally). But how does the failure of a science of consciousness follow from this? It *is* true that there is more to understanding consciousness than understanding the neural facts it super-

venes on. But a complete understanding of consciousness need not be the, or even *a*, goal of a science of consciousness.

Mysterians want to have consciousness explained. A good scientific model of consciousness might go some way toward explaining consciousness, but this isn't an explanation in and of itself. Mysterians pick up on this lacuna and criticize naturalists for promising more than they can deliver.

Mysterians are right that a good scientific model of consciousness won't be an explanation and that the naturalists are confused about this point. However, the mysterians are wrong to expect scientists to address this issue. Explanations are social creatures that depend heavily upon the cognitive resources of the individuals participating in the enterprise: what is explanatory, indeed, intuitive, to one person might be hocus pocus to the next. Some scientific theories can be used in satisfactory explanations because the people involved have the appropriate conceptual frameworks in which to embed the theory as an explanation. But other scientific theories do not have this property. This isn't a problem for the scientific theory, but for the scientists and others who will use the theory. Scientists aren't responsible for the psychological structure of humans; hence, they aren't responsible for the fact that a scientific theory of consciousness will not be explanatory.

There are other nonexplanatory but robust sciences. Consider quantum mechanics. This suite of theories asserts the existence of many things quite counter to our intuitions, including action at a distance and acausal interactions. Nevertheless, quantum mechanics does an excellent job of predicting the future behavior of particles and modeling their past activity. And it is crucially involved in macro explanations such as how chlorophyll works. However, it does not give us a reason why the spin of one particle should be suddenly correlated with the spin of another, widely separated, particle. They just are; that is the way very tiny things work. Quantum mechanics theorizes without explaining – at least without providing us with an intuitive, satisfying explanation. And unlike the case with, e.g., the theory of evolution, quantum mechanics being non-explanatory seems to be a very long-lived property of the science. Whereas evolutionary explanations now strike us as explanatory and satisfactory, quantum mechanics does not and it probably won't for the foreseeable future.

But it is not as though quantum mechanical models are somehow incomplete. They aren't. They do exactly what they are supposed to do, abstract and then model the phenomena. But they don't flow from our physical intuitions, even our well-educated, thoughtful, physical intuitions. Our intuitions derive from wandering around lots of mid-sized objects. Among other things, we

think in terms of causes (*per accidens*, often). As David Hume reminds us, we cannot help but do so, even if in the end there is no such thing as a cause. He is right: if we can't somehow trace out a (*per accidens*) causal chain and grasp how each link is connected to the one before it and the one after, then we don't feel as though we understand what is before us. We need causes in our explanations of change. If we don't have any causes, then we can theorize all we want, but we aren't explaining anything in a psychologically satisfactory way. Essentially the same point can be made for causes *per se*. And this is also quite relevant to quantum mechanics. The hierarchy of processes (or "virtual machines") making up our ordinary world allows us to understand how, e.g., moving cars supervene on working car engines. But this causal hierarchy vanishes when we cross over into the quantum realm. Without such causes, then we can theorize all we want, even very successfully, without explaining anything, for we won't be able to see the supervening behavior as (*per se*) caused by its supervenience base. (Decoherence is thought by many physicists to help here, but this process is not universally agreed upon as the proper explanation.)

Some naturalists have countered that, if given enough time, our intuitions will change to fit the scientific model. The Churchlands are famous for this response. And it could be, we suppose, that simply not enough time as elapsed for our intuitions to shift. However, it is telling that there are now several generations of folk who have grown up with quantum mechanics and feel quite comfortable moving among its hypotheses and models yet still feel the need to indulge in causal reasoning (for example, when their particle accelerators break down). Instead of what the Churchlands suggest, it appears that we just stopped worrying about explaining quantum mechanical "interactions." We accept the models as they are and for what they can do, and we move on to more productive activities. For the most part, we don't sit around wringing our hands about the explanatory status of quantum mechanics. Certainly physicists do not.

We predict something similar will occur with a theory of consciousness. We aren't going to get explanatory scientific theories of consciousness. That doesn't mean that we won't get scientific theories at all – we might get some darn good ones. (Then again, we might not, see the next chapter.) But they won't be psychologically satisfying for exactly the same reason that quantum mechanical theories aren't satisfying – they don't tie into our conceptual frameworks in the right way. Indeed, as we have discussed, they can't.[4]

A science of consciousness is possible, but this science will likely not tell us what it is like to be a bat. But that doesn't prevent it from being science. And, while there is more to understanding consciousness than understanding the neural facts it supervenes on, this, too, has no dire consequences for the exis-

tence of a science of consciousness. Both of these facts are due to an important property of the science of consciousness: it won't close the epistemic gap. This makes it different from many other sciences, but, importantly, not all of them. And, it certainly doesn't prevent it from being a science in good standing.

4.3 More problems with mysterian arguments

Everyone agrees that there are some things we can only know via experience. Though interesting, this is obvious once it is pointed out. Indeed it approaches being a truism: we can only know experience experientially. But are there any things relevant to a *science of consciousness* that we can only know experientially? That's the question. And if the answer is "Yes", then mysterians might be right that our desires for such a science will be thwarted.

According to Diane Raffman (1993), we can know small variations in pitch (those less than chromatic semi or quarter tones) only by experiencing the tones themselves, for no description that we have is adequate. She and others argue that our language-based, descriptive categories are too broad to capture some of our finer-grained perceptual experiences. Raffman's point, like Nagel's and Jackson's, is that there are indeed things that we can only know by experience. Some conclude from such arguments that we have two different bases of knowledge – descriptive and experiential – and without both one cannot have adequate knowledge of perceptual stimuli, nor even, indeed, an adequate account of behavior (e.g., Hubbard 1996).

Though many dual-bases theorists are not mysterians (they are naturalists, mistakenly concluding that, armed with both bases of knowledge, an explanatory theory of the mind, including consciousness is possible), nevertheless, it is easy to turn such notions into an argument for mysterianism. This argument is very similar to the Arguments 1 and 2, above, the only difference being one of emphasis: whereas Arguments 1 and 2 located the central problem in *getting* the relevant information (first-person experience is inaccessible), this argument locates the central problem in expressing it, i.e., it locates the problem in the nature of science itself. Accordingly, it suffers from a different flaw.

Argument 4

1. To be objective, a scientific account of consciousness is going to have to rely on purely descriptive knowledge. This is because a science of consciousness, like any science, has to *be* purely descriptive: one cannot insist that experiencing red be a part of a theory of experiencing red because people might experience red differently (consider someone with an inverted color spectrum – they see violet where you see red; this mere possibility is enough to bar experience from being a part of a scientific theory).
2. Purely descriptive knowledge is not up to the task of supplying a science of consciousness with all that it needs to be adequate. (This is Raffman's point.)
3. So no adequate science of consciousness is in cards.

There are actually three flaws in this argument. The major one is the third. First though, there is the, by now, tired assumption that a science of consciousness has to be explanatory. This assumption is hidden in the use of the word "adequate." Describing a consciousness experience will not provide anyone with an explanation of that experience.

Secondly, positions such as Raffman's are self-defeating. In order for her claim to work, Raffman has to describe the small variations in pitch. Indeed we just did this when we used the phrase: "those less than chromatic semi or quarter tones." (Or, we could chart the tonal distinctions externally, on a computer, for instance, which correspond to the distinctions in our perceptions.) This logical *faux pas* is endemic to any variation on Raffman's point: to make it, one has to supply the very thing that one is claiming can't be supplied (note how Nagel's and Jackson's arguments avoid this problem; they used imagination to successfully point the reader in the right direction). Saying that descriptions are not up to the task required by science by using descriptions is logically hopeless.

The argument seems to work because it is using the well-known property that all experiences have: a description of an experience isn't the experience itself. In truth, no description of middle C is adequate in the sense that you cannot experience middle C via description. But so what? Any argument using this point would simply be a *non sequitur*. That no description of middle C is adequate to experiencing C cannot be a bar to science unless it is assumed that only *experienced* experiences are fair explananda for a science of consciousness. But this is far too strong. The explananda for a science of consciousness are simply experiences, which can be felt or just pointed to via description. The theories of a science of consciousness needn't require that the experiences referred to be experienced. It might be true that a complete and full understanding of hearing middle C requires both hearing it and understanding the theory of hearing middle C. All that follows from this, however, is that a scientific theory of consciousness needn't supply complete and full understanding of hearing middle C. This might seem to be a particular failing of the science of

consciousness, but no science provides a complete and full understanding of its explananda: all sciences abstract and idealize. Nothing in what Raffman argues prevents us from modeling consciousness in as much detail as we would want.

Leaving something out would not be unique to a science of consciousness. There are many objects and relations in physics, for example, for which we have no intuitive, comfortable understanding – something is left out. Examples include: 11 dimension manifolds, wavicles, collapsing wave functions, curved space, quantum entanglement, the beginning of space-time. These are all things for which science has an abstract, highly technical, description but that we cannot understand apart from that abstract description. Intuitively, such things remain mysterious or at least puzzling. Science has no obligation to help us feel an intuitive pull in its theories. What we find intuitive depends upon our other cognitive capacities. Some of the time – a lot of times – scientific theories cohere with our other mental frameworks, but some of time, they don't and can't. In such cases, we still have the sciences, we just lack any intuitive feel for what the theories describe, usually because what they describe is far beyond our ordinary world of tables and chairs, and other human beings.

We conclude that pure descriptions *are* up to the task of supplying a science of consciousness with all that it needs to be adequate. We just change the definition of "adequate" from "explanatory and full" to "predictive and correlational." So even on their own terms, mysterian arguments relying on a fundamental difference between experience and descriptions won't work.

But the truth is, the relevant difference being relied on is the root problem. We have allowed the tacit mysterian assumption that descriptive knowledge exists *apart* from experiential knowledge. That is, we have allowed the assumption that it is possible to have an experience without some sort of accompanying description, and vice versa. Nothing makes two things seem separate as one existing without the other. But, descriptions are never unaccompanied by experiences; they are neither abstract nor general in the ways many assume. And experiences are never unaccompanied by descriptions. We now turn to this.

4.4 Handling a tacit mysterian assumption: The relation between description and experience

We can hear someone speaking to us and understand what that person is saying. We can imagine someone (or ourselves) speaking and understand what is being said. In both cases, our experiences are modality-specific; they are auditory. Whenever we try to entertain some abstract description, we have some

kind of attending experience, usually, either an auditory or a visual experience of some sort. There aren't any purely amodal occurrent descriptions understood in our heads (cf., Barsalou 1999). Descriptions are always accompanied by some particular sensory experience or other.

There are no pure experiences either. Even an experience of the ineffable activates some concepts that allow us to describe the experience. Even if only the concept "ineffable," allowing us to say: "My experience was ineffable." And no experience is purely general. At the least, all experiences are experienced from a certain perspective. (This is the contrary of Daniel Dennett's view in *Consciousness Explained*, by the way. There, he claims that our experiences are not particular and that they are accompanied by general descriptions that give the experiences meaning.)

All we have are semantically informed experiences of the world, whether they be linguistic or otherwise. We parse the world into objects; we can't help but to do so. Contrary to what the positivists wanted, there simply aren't sense-data out there apart from our interpretations of those data. We can't help but see the Necker cubes open up or down; we can't help but understand sounds in our language; we automatically see appropriately arranged marks as words. When we entertain what we know, we are conscious of that knowledge and it has a particular quale in a particular sensory modality.

We are not saying that experience and descriptions are identical or that there is no difference between them. Experiential knowledge is knowledge of some sensory experience as presented in that modality. It underlies our subjective points of view, though all subjective points of view have attached to them descriptive components. Descriptive knowledge is either knowledge of some facts we have no sensory apparatus to detect but is accompanied by some other sort of experience, or knowledge of some sensory experience presented in a different modality. We know what red is like experientially; we have experienced red (a prototypical visual sensation) visually. We know what the Big Bang was like descriptively; we have read and remembered linguistic descriptions of it.

However, we are saying that experience and descriptions don't pull apart the way many mysterians think they do. There is no such thing as purely descriptive knowledge; experience and description occur together. Ultimately, it is best to speak of our human knowledge as having different components, experiential and descriptive, and we use these different components in constructing our points of view.

Scientific theories of consciousness should not differ along most dimensions from any other psychological (or neuropsychological) theory (with the exception of being nonexplanatory, of course). Suppose that our arguments

are correct and that all of our knowledge has both experiential and descriptive components; all our semantic meaning comes packaged with some perception or other, and all our perceptions have some content attached to them. In this case, we can't truly isolate the qualitative experience from its meaning (except in an illusory way) – part and parcel of the *feel* of something for us is its *meaning*. Hence, if we learn anything at all about how we represent the world in our heads, we also learn something about our conscious experiences at the same time. At least this seems a rational hope.[5]

It is a mistake to assume that all of our psychological theories dealing with imagery construction, or sensory perception, or explicit memory, and so on, leave consciousness out entirely. They do not. Insofar as consciousness is an integral part of some of our images or perceptions or memories, any information about any of their properties also thereby gives us some modicum of predictive insight and control into our conscious experiences.

It is also a mistake to think that we will have our biological and psychological theories on the one hand and that, quite separately, we should have a theory of consciousness. If we are going to get a theory of consciousness, it will be intermingled with all the rest of our theories.

Mysterians lament that this view of what a theory of consciousness will look like leaves out the richness of our qualia. Our qualitative experiences are full of details; they are concrete and specific. Any theory of consciousness will not be. This is the point behind Nagel's and Jackson's arguments. It is like something in particular to see red or to track sonar. Abstract that away and you've abstracted out consciousness, the very thing you are aiming to explain.

However, and to say it again, mysterians are operating with a mistaken impression of what a scientific theory is. In order to get useful theories, science has to work above the particular. Scientists theorize about types of events, not event tokens. To understand something, we have to capture what several different instances of that something all have in common and account for that commonality. Of necessity, many, if not most, details will be lost. We take our complex, detail-rich observations and distill them until we are left with their basic structure or their essence or their defining features. We can think of a scientific theory as an abstract picture of an idealized version of our world. We can use this theory to make predictions about what will happen in the idealized version, given certain events under certain conditions. We can then take these predictions back to the real world in all its messy complexity and see how our theory fared. Of course, in the case of consciousness, working above the particular is working above the level that mysterians claim is most important. That may be. But this is a fact of life they will have to learn to live with.

All sciences operate in this fashion. In mechanics, for example, we abstract away from real objects until we are left with point-masses traveling in a frictionless space. We then make predictions about how these points would behave under various conditions. We use this information to explain what happens to objects like cars awash with various frictions. In neurophysiology, we abstract away from all the messy details surrounding ionic influx and efflux in neurons to explain equilibrium in terms of the movement of potassium and calcium ions across the cell membrane. Neuroscientists pick what they think are the essential and basic components of neuronal firing activity and use them (and only them) to account for actual neuronal behavior in all its tangled intricacy. The Nernst model of cell equilibrium leaves out several types of ionic flow; Chomskian linguistics leaves out colloquialisms; Skinnarian S-R behaviorism leaves out the internal processing; the law of gravitation leaves out the shape of the objects.

Psychology functions the same way. For example, Stephen Kosslyn's (1980, 1994) distinction between surface and deep representations as the two major components of visual imagery abstracts away from other important components of visual processing. Feature binding, grouping effects, lighting assumptions, and so on, are ignored. This is not to say that Kosslyn believes that our brains do not do these things. Nor does this mean that Kosslyn thinks that these processes are unimportant. Kosslyn's model focuses on a few features that he takes to be extremely salient. Another psychological model, with a different emphasis, would take other things as its defining features.

Any scientific theory of consciousness should work in the same way. Since individual conscious experiences are personal and particular and since the personal and particular are not under the purview of science, we can and should abstract away from them. Unlike in most other sciences, this abstracting is quite expensive when developing a science of consciousness (for it leaves out what we intuitively think ought to be explained), but it is a cost that has to be paid in any case (this the conclusion of Part 1). A good theory of consciousness, then, will abstract away from all the specific details of our phenomenological experiences and pick out a few defining features, not because these details are unimportant, but rather because we can't theorize about them as details.

It is wrong to complain that Kosslyn's model doesn't explain why our visual images correspond to perceptual experiences. That is to say, theories of imagery don't explain why our images and our perceptions use the same parts of the brain in more or less the same way. That they do is something we discovered and that we now (should) take as our starting point in theorizing. Kosslyn builds models of how our imagery *qua* "rewritten" perceptual experi-

ence works. He, and others, articulate what it is in our brains that corresponds to our images or our perceptions. He elucidates and clarifies the processes that terminate in our visual experiences. He refines and regiments the components of these experiences. He does not, however, *explain why* the visual images and visual perceptions occupy similar spaces in our brains. It is just the way we are. (An evolutionary story might answer these questions, but this isn't what the mysterians are looking for either.)

A scientific theory of consciousness will work in the same way. We might discover that our conscious experiences have certain components, which, once discovered, scientists can take as starting points in theorizing about our experiences. They can then try to build models that capture how our conscious experiences qua these basic facts function or point to aspects of our brains likely to be instantiating these facts.

Mysterians are not happy with this position. Unfortunately, the only way to remedy this situation is for them to give up on their impossible demand.

Chapter 5

Science in the face of mystery

> ... it will not have escaped the notice of those interested in the topic that we have, at present, nothing resembling a science of consciousness.
>
> *Anthony Jack and Tim Shallice*

5.1 The science of consciousness in broad outline

The mysterian argument gets off the ground by assuming the need to intimately relate, via reduction, two things: the subjective and the objective, the phenomenal and the physical. We have argued that this is misguided: neither full translation nor any sort of reduction is required for developing a useful, predictively adequate theory of consciousness.

But, if the science of consciousness is not going to explain consciousness, what will it do? Taxonomize, correlate, predict, and control primarily. It will work out detailed taxonomies of neural processes and kinds of conscious states, including perceptual states, emotions, consciousnesses associated with propositional attitudes, and perhaps even such phenomenal states as experiences of the mystical or the divine. It will map correlations between these various neural states and neural processes at various levels and corresponding conscious states. It will make predictions. These predictions will take the form of "if we do such and such to this neural state (or subpart of a neural state), the subject will experience E." and "if the subject experiences E, then the subject's neural processes should do such and such." How will consciousness scientists know that the subject is experiencing E? By its behavior, either by asking the subject or detailing its nonverbal behavior. And we can use all this information to control conscious states, either our own or those of others. (Indeed, we are well on our way in this last venture, with all the mood altering drugs on the market these days.)

This description of the science of consciousness might sound thin, and it is thin to the extent that sciences are required to produce satisfying explanations. But our plaint all along has been that this requirement is too strict. Scientists

are very pragmatic. If reductive explanations are not in the cards, then we get taxonomies, correlations, and predictions instead.

There is already a science of consciousness, toiling in the fields, comprising cognitive psychology, cognitive neuroscience and psychophysics, and all of these are replete with correlations. This is not an indication of the imminent failure of the science, but instead it is its most promising feature. Of course, we cannot predict how this science will unfold. All the relevant information is in the future. But we can say what it *won't* have to do. It won't have to explain consciousness; it especially won't have to explain it in a way that satisfies anyone's intuitions.

5.2 Overconfidence, underdetermination, and the correlates of consciousness

But no matter how one comes down in this debate, it is true for all concerned that science is *at least* in the business of seeking the *correlates* of consciousness. As it is now fashionable, we will refer to the correlates of consciousness as the NCC for "neural correlates of consciousness" (though our remarks below also extend to any alleged psychological or microphysical correlate as well).[1] At the least, science is in the business of finding physical differences between someone who is conscious and someone who is not. Seeking such a correlate is a reasonable goal, even if one is a hard-core mysterian.

Still, uncovering the NCC is not as straightforward as it may seem, for it is a nontrivial question which correlate should count as *the* correlate. It is not an empirical difficulty we are alluding to here, though obviously there is that, too. We are concerned with a deeper and prior methodological or philosophical question: How do we pick out the correlate from boundless other spurious coincidences, given that we have to draw our conclusion from spotty and profoundly indirect evidence? This is a question about the appropriate level of analysis in the brain for understanding and explaining consciousness, and a question about how to make convincing cases for inductive inferences using less than crystal clear data. Though this question exists in some form in all of experimental science, it is unusually sharp and difficult when the issue is consciousness. As a result, we are even further away from a theory of consciousness than one might think (even an non-explanatory theory).

To explore this issue, we begin by taking as our stalking horse Hans Flohr's identification of the correlates with the NMDA receptors in the cortex, for this is a particularly rich and well-developed proposal. We note that there are others, including Bernard Baars and James Newman's reticular formation, Francis Crick and Christof Koch's (and others') 40 Hz oscillations, Hardcastle's activation in the parietal cortex, and Stuart Hammeroff's microtubules (Baars & Newman 1994; Crick & Koch 1990; Hardcastle 1995; Hammeroff 1994). The worries we raise here about actually uncovering the NCC are equally applicable to any and all of these; in fact, they apply to any member on our original list in Chapter 1.

5.2.1 Flohr's hypothesis[2]

Flohr claims David Hebb was right (Hebb 1949): brains *are* plastic. Any activity between synapses strengthens the connections, so post-synaptic neurons fire more readily the next time around. Inactivity weakens synaptic connections, so post-synaptic cells are more difficult to provoke into firing later. The brain's mantra is "use it or lose it." This so-called Hebbian learning rule means that brains will develop complexes of neurons which prefer to fire together when a subset of them are stimulated. These assemblies, say Flohr and Hebb, are the building blocks for mental representations. In other words, these groups of cells that prefer to fire together as a unit are the neural correlates of our mental representations.

Unlike Hebb, though, Flohr, along with Christof von der Malsburg (1981), holds that there are two ways to create cell assemblies. On the one hand, we can see permanent change occurring slowly over a period of time and repeated activations. As we experience and learn about our environment, our neural connections shift and grow such that we develop privileged neural units that fire in response to particular environmental stimuli. On the other hand, transient assemblies are also possible, and these occur quite rapidly, on the order of 100 msecs. These are assemblies that are not carved out by repeated experience; they are not learned. Instead, they are formed spontaneously, on the fly, as it were, in response to whatever else is going on in the brain at that time.

Flohr's research has indicated that particular neural receptors in the cortex are responsible for both types of changes. These receptors, known as the NMDA receptors, are well known in neuroscience as one of the receptors found in the synapses of some neurons that support the changes in the brain due to learning ("NMDA" stands for N-methyl-D-aspartate). Basically, they work to increase the connection strength between two neurons if those neurons are stimulated together by increasing the neurons' sensitivity to incoming signals. NMDA receptors are likely the mechanism by which Hebbian cell assemblies are formed. If, as some researchers think, changes in connective strength by

the NMDA receptors comprise the cell assemblies that underwrite representation in the brain, then these receptors are the mechanism by which mental representations (or their neural correlates) are formed.

According to Flohr, some of the rapid, transient cell assemblies are conscious. Conscious representations, Flohr thinks, are those assemblies, those (correlates of) mental representations, which are *self-reflexive*. These are the cell assemblies that refer to themselves as referring to something else. Less cryptically, these cell assemblies refer to their current state in the brain as well as to something external to the brain (or to themselves). According to Flohr, being a system that has self-reflexive cell assemblies is a necessary condition for having phenomenal experiences.

A second neurophysiological item important for consciousness, according to Flohr, is the ascending reticular activating system. This system is a large series of neural tracts originating in the more primitive parts of our brain near our spine and then spreading throughout the cortex. It doesn't appear to be activated by any particular modality of input; hence, it is referred to as an *unspecific activation system*. It seems just to boost brain activity generally.

Bilateral lesions to the reticular formation lead to deep coma, and, Flohr presumes, unconsciousness. Somehow, the thought goes, the reticular formation interacts with specific transient cell assemblies to produce, enhance, or modify consciousness. Flohr hypothesizes that this system determines how likely it is that a cell assembly forms as well as aids in binding together several simple assemblies into more complex representational states. If cell assemblies are created quickly enough, then the system is conscious: "an unconscious state is present if the rate at which plastic changes take place falls below a critical threshold" (Flohr 1995b: 160). So, according to Flohr, consciousness is connected to rapidly formed cell assemblies. Cell assemblies form rapidly through the activity of the NMDA receptors and the ascending reticular activation system.

In his theory, Flohr outlines lots of different events in the brain that co-vary with consciousness: self-reflexive representations, rapidly changing cell assemblies, activation in the reticular formation, NMDA receptor-driven synaptic change. It strikes us that all are possible candidates for *the* neural correlates. Flohr himself, however, has no difficulty in choosing the one item he believes is crucially responsible for consciousness: "The occurrence of states of consciousness critically depends on a specific class of computational processes that are mediated by the NMDA synapse." Because "the direct or indirect disruption of NMDA-dependent processes is the common operative mechanism of anesthetic action," he concludes, "the essential difference between anesthetized brains and conscious brains consists in the presence or absence of NMDA-dependent computational processes" (Flohr et al. 1998). According to Flohr, NMDA-sponsored computations are the neural correlates we all have been looking for.

Let us leave aside some of the potential difficulties of this view – for example, why Flohr wants to claim that NMDA synapses are crucial for consciousness, even though their activity is just as essential for non-conscious cell assemblies (the non-reflexive ones, or the ones which do not fluctuate rapidly enough, or the ones we find in sea slugs). Let us just assume that Flohr's theory is correct in its entirety: A phenomenological experience is a reflexive cognitive thought, which, neurophysiologically speaking, is a complex and transient cell assembly in the cortex, underwritten by activity from the reticular formation, and made possible by the NMDA receptors and their computational properties.

So which *are* the neural correlates of consciousness? The cell assembly itself, as Hardcastle has argued? The reticular formation, as Baars and Newman claim? The NMDA receptors with their computational properties, as Flohr believes? Or perhaps their internal quantum effects, as Hammeroff holds? How do we know that what Flohr suggests is the neural correlate? All are correlated with the qualia, after all.

5.2.2 Is there a way to find the NCC?

Can creatures with cell assemblies but no reticular activating systems be conscious? How about those with a reticular activating formation but no NMDA receptors? Or NMDA receptors with different computational effects? Or the same computational effects but with subtly different underlying quantum interactions? Just run the right sort of experiment and scientists should be able to separate the genuine effect from mere experimental artifact or interesting coincidence.

5.2.2.1 *Some philosophy of science: The method of screening-off*

The investigative strategy philosophers of science advocate for determining the actual cause of an event is to determine which putative cause "screens off" the others (cf., Cartwright 1970; Eells 1988; Hardcastle 1991, 1998; Salmon 1971; Wimsatt 1984). In general, we say that one putative cause, *A*, *screens off* another, *B*, from the effect we are interested in if the probability of the effect occurring is the same, regardless of whether we have *A* and *B* occurring or just *A*, but the probability decreases if we have just *B*. In other words, having *B*

does not add anything to the system. In such a case, we can conclude that *A* is actually doing the causal work.

A screening-off analysis can be run either horizontally or vertically. That is, we can use it to determine which cause is "closest" to the explanandum and so is the most proximal cause. This is the horizontal direction of screening-off. Or, we can use it to determine which level of organization of matter is the appropriate one for describing the phenomena. This is the vertical level.

To see how the screening-off relation functions, let's look at some examples unrelated to consciousness. First, consider a simple horizontal example: bowling. What causes the bowling pins to fall over, the bowling ball hitting them, or the person rolling the ball? Answer: the ball. The pins fall over if the ball hits them, regardless of whether the person rolls the ball. The ball could have been pushed; it could have been fired out of a cannon; it could have been sent careening accidentally towards the pins by a car crashing through the front of the bowling alley. Certainly, the person throwing the ball is causally relevant, but the proximate cause of the pins falling down is the bowling ball hitting them.

The units of selection controversy in evolutionary biology also can be understood to be an example of horizontal screening off (cf., Brandon 1982). We would say that the phenotype *A* screens off the genotype *B* – and hence is the causally relevant factor – with respect to reproductive success if (1) we can affect reproductive success by changing the phenotype, (2) we can change the phenotype without altering the genotype, and (3) changing just the genotype does not affect reproductive success.

Phenotypes emerge through the interaction of genes, the various kinds of RNA, and the environment. DNA comes first and the phenotype later. Since phenotypes screen off genotypes, we know that phenotypic traits are the most proximate relevant causal factor in reproductive success. That is, phenotypes come later in the developmental chain and so are "more" responsible for the final product than are genotypes. As long as we have the phenotype, we get the units of selection. The same is not always true for the genotype. In general, we use horizontal screening-off relations to outline (per accidens) the casual mechanisms for some event.

The conflict between psychiatric and neurobiological explanations over explaining something like depression illustrates the vertical dimension of a causal analysis. On the one hand, we might want to claim Fred is depressed because he learned that he has cancer. The cognitive event of understanding that one has a potentially fatal disease is the causally relevant factor in explaining Fred's change in mood. On the other hand, we might also want to claim that Fred is depressed because the amount of norepinephrin has dropped in his brain. The

amount of a particular neurotransmitter is the causally relevant factor in explaining Fred's change in mood. Which explanation is the correct one? Which one picks out *the* cause of Fred's depression?

To answer these questions, we need to know whether cognitive events screen off neurophysiological ones, or vice versa. Is the probability greater that Fred will get depressed if he learns that he has cancer or if he learns he has cancer and his norepinephrin level drops? Or, is the probability greater that Fred will get depressed if his norepinephrin level drops, but he hasn't learned anything new? Research indicates that neurotransmitter levels are better indicators of depression – Fred is more likely to get depressed if his norepinephrin level drops than if he receives some bad news – hence, we should look to neurophysiology and biochemistry to understand large scale mood swings. Changes in our brain chemistry screen off psychological events, in this case. We use vertical screening-off to determine correlates, which can be used in some instances to ground reductive explanations.

Discerning the neural correlates of consciousness is also a vertical problem. We want to know which level of organization in the brain is the appropriate one for understanding consciousness. Which events in the brain, the assemblies themselves, NMDA-computations, or microtubule activity, are most closely associated with conscious experiences? Which predict its appearance best?

5.2.2.2 *Accessing consciousness*

Unfortunately, running the right sorts of experiments is easier said than done. The first (and very obvious point) is that we only have indirect access to others' conscious experience. In general, we take verbal reports (or related behaviors) to be by-and-large veridical descriptions of phenomenological experiences. And we take lack of verbal report as evidence for not being conscious. These facts mean that whatever we believe about the presence or absence of consciousness is going to be skewed by our beliefs about language and mnemonic processing, on the one hand, and the subjects' capacities for self-report and memory, on the other. While we all recognize that a subject's saying something doesn't make it so, in practice it is quite difficult to devise experimental tasks that do not explicitly assume that verbal reports index conscious experience; indeed this assumption is usually openly made.

Our reliance on verbal reports of either previous or current experiences, and reporting's dependence on linguistic and mnemonic capacities, means that we cannot be entirely sure when someone or something is conscious. Most researchers hold that anesthesia, stage four sleep, and deep coma render organisms unconscious. These are mere assumptions at best; and dubious at next

worst; and false, at worst. Why would they think this? It is true that one can only rarely and then only with difficulty report any phenomenological experiences when roused from one of those states. However, it is not at all clear that this isn't a problem with memory – we cannot remember what we experience when we are in those states. Or, perhaps it is a problem with linguistic access to these memories – we can remember these states, but we cannot put these memories into words. Since both memory and language mediate our access to others' conscious mental states, we are barred from devising experiments that compare test subjects with phenomenological experiences to non-experiencing controls in a reasonably pure form. Though we may have paradigm instances of conscious experiences, we certainly do not have pure and uncontroversial cases of unconsciousness (except perhaps in the case of death, but then there are too many other confounding features for good experimental practice).

This is important because most consciousness researchers rely on juxtaposing paradigm cases of conscious experience with presumed cases of nonconsciousness (sometimes to certain stimuli) to support their pet hypotheses. But if we cannot be certain that someone is unconscious, then *ipso facto* we are equally unsure of nonconscious states in otherwise conscious subjects (for those states are just unconscious states). Hence, we cannot use such differences to support any ideas concerning what in the brain is associated with phenomenological processing. Surgical patients, for example, can show later emotional responses to events that took place while they were under anesthesia; indeed, they sometimes report conversations occurring in the operating room. People aroused from stage four sleep report perseverative thoughts. Such facts don't inspire confidence in the contrastive data used in neural correlates discussions, and contrastive data accounts for most of the data in consciousness studies.

Bilateral lesions in the reticular formation cause deep coma from which patients cannot be roused. It is reasonable to conclude that the reticular formation is somehow tied to alertness. But what does this fact have to say about conscious experience? Next to nothing, we think, since people in a deep coma cannot report what their experiences are like, nor if they had any in the first place. If we already knew that the reticular formation is the neural correlate we seek, then we could predict that patients with damaged reticular formations have diminished consciousness, but we don't know that yet. Or if we knew that coma and alertness were deeply tied to consciousness (*knew*, not just *suspected* or *commonly assumed*), then we'd have evidence that the reticular formation was tied to consciousness. But we don't know that either.

Certainly, the differences between alert and anesthetized patients drive Flohr's theory. Patients under anesthesia cannot be roused either (which is a good thing, obviously). It is reasonable to conclude that anesthesia is also tied to level of alertness. But, again, what does this fact have to say about conscious experience? Still next to nothing, since patients under anesthesia cannot report what their experiences are like, nor if they had any. Suppose, for example, that anesthesia causes profound amnesia and some sort of bodily immobility, rather than an inability to feel pain. In this case, patients on the operating table would behave exactly as they do now, but would still be conscious. (This isn't a mere thought experiment either, for, in the United States anyway, surgeons add amnesics and paralytics to the anesthesia administered during surgery.)

Whether we think someone is unconscious depends on what they are later able to report. If they can't report anything, we typically believe they were unconscious. Our point is that this inference isn't justified. If what is wanted is certainty, then there just isn't any. If what is wanted is high experimental confidence, then there isn't any of that, either. If we already knew that anesthesia blocks consciousness, then we could predict that patients under anesthesia are unconscious, but we don't know that yet. In short, most evidence touted as relevant to consciousness is still as of yet not evidence at all; it is begging the question. Consequently, designing the appropriate experiments for uncovering the neurophysiological event that screens off the others vertically is at best an extremely difficult task.

5.2.3 Blindsight and other philosophical examples

But, some might protest, we aren't being fair to the wealth of scientific data out there. In particular, there are good neurological examples of cases in which a person demonstratively loses qualitative experience, cases in which the evidence isn't tainted by alternative hypotheses, cases in which cognitive scientists aren't begging the question. Blindsight is the example commonly given by philosophers (and others). Let us take a brief detour to examine this claim in some detail lest it be taken as a way around the difficulties with devising experiments on consciousness.

It was once believed that lesions in the primary visual cortex (area V1) resulted in complete blindness. It seemed reasonable at the time: V1 was supposed to be the terminal information processing station along the primary visual pathway, our only visual processing track. But then, in 1967, Larry Weiskrantz and Alan Cowey reported that monkeys whose striate cortex (which includes V1) had been removed acted as though they could see. They

didn't behave as if they were blind, being unsure of their surroundings or making hesitant movements. They moved fluidly; they picked up objects allegedly in their blind field; they went boldly about their business of living.

As in monkeys, so too in humans. When we lose parts of V1, we complain that we can't see in the visual regions that correspond to the damaged areas. We complain, but we can actually perceive things there, at least up to a point. We behave as though we are seeing; we just aren't *conscious* of our seeing. Mapping out exactly what we can and can't do in blindsighted scotomas has become something of a cottage industry in neuroscience, as are mapping out other oxymoronic states like numb-sense and deaf-hearing.

Philosophers actually anticipated this deficit in blindsight-type thought experiments (cf., Perkins 1971). David Mellor (1977) properly introduced blindsight to the philosophical community when he noted, rightly, that blindsight gives us an actual case in which function (of a limited sort) and experience pull apart. It looks like here, maybe, we can distinguish the phenomenological feel of something from what we are able to do. Theorizing about blindsight and what it means for consciousness studies has become a thriving enterprise in philosophy as well.

To put the discovery of blindsight in the language of screening off, it looks like we have found a case in which our neurophysiology screens off phenomenological experience with respect to some behavioral task. To put the claim in its starkest form: we need our brains to perceive, but we don't need consciousness. We can perceive, even if we lack awareness.

But we know more than this, too, for blindsighted individuals are behaviorally distinguishable from sighted ones. Blindsight patients can only discriminate crude stimuli in their scotoma in a rough and ready way; they can tell (guess better than chance) direction of movement, color, shape, and number. But they can't read. They can't thread a needle. So, maybe, too, we know that we need consciousness in order to process difficult stimuli. Though we don't discuss why here, it also looks like we might need it to initiate intentional behavior. Here then is the idea: for basic perception, we don't need consciousness, we just need the appropriate tracks in the brain; for complex perception and voluntary action, we need consciousness (and the appropriate brain tracks). Find the parts of the brain correlated with such complex perceptions and intentional action, area V1, say, and we will have isolated at least one NCC.

Is this right? If we are patient, will Mother Nature herself tell us what we need to know in order to isolate the NCC?

Let us leave aside the still unresolved question of whether blindsight is a real phenomenon and not an artifact of residual functioning in an incompletely

lesioned striate cortex (cf., Gazzaniga et al. 1994). Even if blindsight were real and functions exactly as we are imagining, we still have a problem – we don't know what we need in order to pick out a unique NCC. And we won't for the foreseeable future.

Blindsighted patients seem unaware of what is in their blindfields, it is true. While it may be that their degraded performance is due to lack of consciousness, it may also be that whatever knocked out consciousness also impairs their performance. Consciousness might not be necessary for optimal perception and action, after all. Zombies might really be possible – in this, the actual world. It might be that consciousness is not directly tied to performance, but only indirectly through some common cause. Losing part of V1 might both affect consciousness and then, independently, affect function.

Bedrock slipping on either side of a fault-line will be followed by a rumbling noise and perhaps pictures falling off your wall. However, does the rumbling noise cause the pictures to fall off your wall, or does the slipping bedrock cause both? We know that it is the earthquake that causes both. The question we need to answer is: Is losing part of V1 like an earthquake, doing more than one thing at a time?

Losing part of V1 as a common cause would screen off both consciousness and function, but consciousness would not screen off function, nor vice versa. To sort out whether this scenario is correct, we would need to isolate consciousness from complex perception and intention, or, if we can't do that, then isolate V1 from complex perception and intention. However, we can't do either of these things. We don't know how. If we knew that consciousness was required for complex visual perceptions, then the blindsight studies would tell us about the processing capacities of consciousness. But we don't know that. That is what we are trying to figure out.

The predicament here is the same as with Flohr's hypothesis. We can't investigate the neural correlates of consciousness without already assuming we have isolated the correlates. Furthermore, as long as we are unable to access conscious phenomena directly, apart from some sort of speech act or other behavior, we shall always run into this difficulty. We might have isolated the correlate, but then again, we might have isolated some common cause that affects both consciousness and the marker we are currently using for consciousness. Without being able to identify consciousness in terms of some objective attribute, we can't get any sort of experimental program off the ground.

5.3 The pragmatics of consciousness research

What should we do? Given the undeniable lack of data, how are we supposed to determine the neural correlate of consciousness? Even if we suppose that something like Flohr's theory is correct, how can we prove that NMDA-computations screen off everything else?

The best answer is that we should turn to the pragmatic aspects of explanation and the various explanatory heuristics science has adopted (for better or ill) over time. William Wimsatt (1984) argues that we should relax our notion of screening off. In actual science, with real world constraints, we perform a cost-benefit analysis so that we would say that *A* "effectively" screens off *B* if adding *B* to our explanation increases our understanding of the effect only a small amount and *B* is difficult or expensive to procure. Determining whether some variable screens off another is partially a pragmatic decision. In all explanations, some otherwise relevant events are set aside as being not significant enough to warrant including. Not all causal influences are created equal, and we need worry only about the most obvious in our explanations and research.

For example, gravity is necessary for human consciousness, for without gravity, life itself would not be possible. On the other hand, when singling out items of study for consciousness in the brain, gravity does not rank high on the list of possible suspects since there are other things also necessary that are closer to being sufficient for consciousness as well. We are searching for the neural correlate that is both necessary and sufficient. Since we can't have that, we want one that is as close to being sufficient as possible.

Let us put this point a different way. When we devise scientific theories to explain phenomena in the world, we single out some causal relations among the vast web of interactions as the important connections for understanding some event. Some causal influences are too trivial to matter much in a general theory. Others are important, but too far removed from the phenomenon in question to fall within the scope of a finite theory. Under normal circumstances, UV radiation would fall in the former category for any general theory of consciousness and breathable air would fall into the latter. Though one's skull is constantly bombarded by ultraviolet radiation, its effects on one's conscious experiences are quite minimal. Hence even a complete theory of consciousness can ignore UV effects, for successful theories highlight only the most important components and interactions in getting to the specified end state. Oxygen supports mammalian life and being alive is a prerequisite for our being conscious (probably). However, that we and creatures like us are alive and consuming oxygen are not going to be facts included in a theory explaining our conscious awareness. We can take life sustained in part by oxygen as a given background assumption and build a theory of consciousness on top of such facts.

Which neural correlate we chose will be at least partially determined by which gives us the most bang for the intellectual buck, as it were. This concession to the social pressures on science goes a bit of the way toward answering the worry. We need not consider the conscious experiences of creatures with brains very different from our own when theorizing, since we have little-to-no access to their mental lives – what is it like to be a bat, indeed – and, more importantly, such data would be very difficult to get. To explain our own consciousness under normal conditions to other like-minded creatures would be enough. We need not include bats or computers or Martians in our scientific conversations.

However, even though the pragmatic aspects of theorizing are important, they will not completely solve our central concern: the vertical question of which event, at which level of organization, in the human brain is most closely associated with consciousness. In short, we still can't find the correlate. Data that separate NMDA-receptor computations from their quantum effects or the formation of cell assemblies are, at least for the moment, impossible to get for intact brains. *Ex hypothesi*, they are all perfectly correlated with conscious experience, and higher-level events are determined by lower level events. We really can't get one without the other. Is determining the neural correlate, then, an insurmountable problem?

We have one explanatory move left to us at this point, and that is to turn to previously accepted explanatory heuristics in science: simpler is better; reduce as far as possible; consiliance and parsimony are preferred; and so forth. These set the standards for ideal explanations. The best explanation of some phenomenon is one that is simple to model, contains few variables, and dovetails nicely with other previously accepted theories. These sorts of explanatory goals inform scientists' hypothesizing. No one is going to propose, much less have accepted, a complicated theory if there is a simpler one available.

In addition, in biology and neuroscience, there has previously been a distinct bias toward reductionism. The assumption is that the smaller the unit of analysis, the better, for the more fundamental processes occur at the lower levels of organization. With this bias in place, we should say that the neural correlate of consciousness is most likely the smallest neural unit we can discern that co-varies with consciousness. In this case, someone like Flohr or Hammeroff would be right: the neural correlate of consciousness is probably something like receptor computations or quantum effects in microtubules.

Recently, however, this propensity toward "smallism"[3] has come under fire. With the increasing popularity of large-scale dynamical systems explanations of brain phenomena, we are all losing our unspoken agreement that the real stuff occurs down below, while the surface appearances are mere reflections of the underlying causal interactions. Nowadays, we are finding champions of "largism" at every turn (cf. Kelso 1995; Port & Van Gelder 1995; Skarda & Freeman 1990; Thelen & Smith 1994). These researchers disdain the small as relatively irrelevant data and seek true understanding in the large-scale patterns that emerge out of the chaos of tiny interactions, each of which is insignificant when considered alone. A largist would claim that the complex cell assemblies are the true neural correlate of consciousness, while the microtubules and NMDA-receptors merely support the assemblies.

Which way should we go? Which explanatory bias should we adopt in consciousness studies? Unfortunately, the answer is not forthcoming. We are caught in an odd time, as a war over explanatory biases in biology and neuroscience is just now being fought in journals and laboratories around the world. Some neurobiological explanations are strongly reductionistic; others are not. And neither have the upper hand at the moment with respect to explanatory power, funding decisions, centrality in the profession, and the like (cf., Hardcastle 1998).

Hence, we cannot say with any sort of surety what the correlate of consciousness is. Whether we should be smallists or largists in our explanations of mental events is currently undetermined. And we have too many levels. For now, we have only educated guesses, personal declarations of faith, and a plethora of individual research programs. Lots of basic research remains to be done, and, more importantly for our concerns, the fundamental theoretical scaffolding remains to be constructed. For now, finding the correlate of consciousness is itself a deeply difficult problem.

5.4 The naturalists' promissory notes

It might strike you that there is something profoundly dissatisfying about this discussion. We want to know what consciousness *is*, what in the brain (or wherever) causes consciousness to be. It doesn't seem that current funding fads at the National Science Foundation should determine what that answer is. It just seems wrong that serious ontological questions turn out to be merely pragmatic decisions.

This conclusion is especially troubling given that there has historically been little consistency in the mind/brain sciences regarding what sort of data count as the right sort in answering ontological questions of the mind. At the turn of the century, when introspectionism reigned in psychology, first-person phenomenology was privileged. Then, with the advent of logical empiricism, behavioral evidence ruled and phenomenology was considered to be something of an embarrassment. In the late 1970s and early 1980s, data from brain chemistry trumped when eliminative materialism had its fifteen minutes of fame. Maybe next week, the tide will turn again. . . . No wonder the mysterians have abandoned hope.

The shifting sands of explanatory bias in the mind/brain sciences drive home how little one gets out of a scientific theory if one is fundamentally interested in ontology. What scientists will end up identifying with consciousness will depend on how other, largely unrelated, research programs fare. This isn't a criticism of the science; there is simply no other way to run the business. But it does give us a sense of how large the promissory notes the naturalists are issuing, and how wrong-headed.

Yet, the project did seem so reasonable – mundane, even. Consciousness has to be correlated with something in the brain. Isolate that something and, bingo, consciousness has been identified. But, once we start to push on the naturalists' program, it falls apart in our hands. We can't isolate consciousness because too many things are correlated with it at too many different levels of organization. And we can't even be sure of the correlations because we only have indirect access to other peoples' first-person experiences, to their consciousness. We are forced to turn to other factors to help winnow our data. These social, cultural, historical – *prudential* – factors aid in theory-building. But they shouldn't help in our understanding of fundamental ontology; they are simply the wrong sorts of beasts to do so.

Consider: from a materialist's perspective, being able to taxonomize, correlate, predict, and control is enough to be able to identify conscious phenomena with physical things in the world. From a dualist's perspective, being able to taxonomize, correlate, predict, and, control is enough to point to some established harmonies between the physical world and the non-material one. However, because the actual theories advanced by these two camps' science would not differ, one should be indifferent between which metaphysical option one chooses. The science remains the same. The ontological difficulties drop away.

This is not to say that there is some principled reason preventing scientists from working out *some sort of* theory of consciousness. Very likely, at some

point down the road, scientists will produce a useful theory; we don't want to second-guess what scientists can accomplish. But it is to say that in spite of consciousness's strange properties, the science of consciousness, like any other science, is a social process, governed by social conventions, and utilized by socialized creatures. And it is also to say something further: *because* of consciousness's strange properties, the science of consciousness is governed more than other sciences by stipulation and pragmatics. As such, it has philosophical, especially epistemological, limitations.[4]

This difficulty is peculiar to consciousness. With other sorts of inaccessible phenomena, we can still garner converging evidence that suggests their etiology. We hypothesize that dinosaurs were living animals because their fossil record is homologous to the bones of other animals, for example. But we can't do this with consciousness. We can't get converging evidence for its correlates. We have no way of accessing consciousness, except by already agreed upon markers and correlates. But these are exactly what we are trying to discover. So we are blocked. And no experiment can save us.

But, if we gave up on searching for *the* NCC and were content with a set of "CC"'s – whether they were the quantum effects of neural microtubules or populations of neurons and gross neural processes or psychological markers – then we could skip around this logical roadblock. As long as such correlates were useful for other theorizing and other sciences, then they should all be admitted to the club of sanctioned scientific entities.

To make a broad generalization, sciences relate things. They relate (types of) properties, primarily. Broadly speaking, science produces two kinds of relations: (1) relations from higher level facts to lower level facts, and (2) relations between facts couched at one level. Statistical thermodynamics and genetic explanations of phenotypes are well-known cases of the former, while Boyle's law and the pendulum law are examples of the latter. Sciences of both types can offer reductive or causal explanations, but they need not.

The science of consciousness will be a maverick member of the scientific community. However, like all other sciences, the science of consciousness will relate one set of facts (those about consciousness) with another set of facts (those about something else). Basically, the science will map and exploit correlations between the phenomenal realm and the neural or psychological realm. It will offer what we dub *nonprivileged pure correlations* between conscious states (hopefully taxonomized in certain ways) and psychological or neural or quantum states, again suitably taxonomized. Pure correlations are *explanation-free*. Pure correlations tell no stories to connect the lower level to the higher one, conceptually. Nonprivileged correlations refer to a set of correlations in

which we cannot privilege any particular one as being more fundamental for our theories than any other.

Nevertheless, evidence for nonprivileged pure correlations can be quite strong, and confidence in the correlations can be quite high. Confidence is a function of the robustness of the taxonomies and the predictive adequacy of the science of consciousness's predictions.

Part III

An application

Consciousness and philosophy

Chapter 6

How consciousness creates philosophy

> That hour, like a breathing-space which returns as surely as his suffering, that is the hour of consciousness. At each of those moments when he leaves the heights and gradually sinks toward the lairs of the gods, he is superior to his fate. He is stronger than his rock.
>
> *Albert Camus*

We have argued that the classic problem of consciousness – the problem of *explaining* in some satisfying way how consciousness supervenes on brain processes – is unsolvable because it is impossible to be sure whether consciousness logically supervenes; indeed it is conceptually impossible to understand how consciousness *could* supervene. Yet, for all that, consciousness could logically supervene. Then again, it might not. Hence, naturalists are wrong. However, the problem's insolvability does *not* mean that there will be no robust, scientific theory of consciousness. There probably will be. It won't be a reductive, explanatory theory, but it will be a theory, nevertheless, and will traffic in non-privileged pure correlations. Hence, mysterians are wrong. Unfortunately, and continuing the negative news, the "probably" here must be taken seriously, since, as we also argued, the antecedent probability of successfully developing a science of consciousness is lowered by the fact that finding NCC's is going to be especially difficult. But whether there is a theory of consciousness or not, scientists can and will study consciousness, which is some consolation.

It is instructive to compare our view to a dualist's view like that of Chalmers. Like us, Chalmers argues there will be no reductive (satisfying) explanation of consciousness, yet there will still be a science of consciousness based on its naturally supervening on the neural level. But this is a surface similarity only. For Chalmers, it is a *metaphysical fact* that consciousness doesn't reduce to or logically supervene on neural processes. The science he sees in the offing is based on bridging principles local to our possible world between the phenomenal realm (which he claims exists) and the physical realm (1996). Our view is that no such metaphysical fact can be established. For all intents and purposes, there is *no* such fact. We think that the epistemic situation dominates, indeed constitutes, the entire landscape. We cannot know that dualism

is true. And we cannot know that materialism is true. The reason that there will be a science of consciousness is not due to bridging principles, but rather due to the fact that it is possible to have science without reduction. In place of bridging principles, there will be (nonprivileged pure) correlations. These correlations might bridge two disparate realms, the phenomenal and the physical, but they also might relate two physical domains, the conscious and the neural, that we are incapable of conceptually linking because of the nature of phenomenological experience itself. If being conscious creates the *illusion* of consciousness's nonsupervenience, then materialism could well be correct. Our position is that we can never dispel the possibility that dualism is an illusion. Hence, no conclusions can be drawn from conceiving of dualism.

It is a consequence of our view that the debate between dualism and materialism should be abandoned. It is unsolvable for principled, logical reasons. Another consequence is that, ultimately, we will have to be liberal about what we count as a science of consciousness. When science confronts consciousness, it is science that changes.

Our position is disconcerting perhaps, but it does look like the right view of things. Given this, consciousness researchers should give up any desire for the science of consciousness to traffic in satisfying, reductive or bridging explanations. That is, we recommend abandoning desires for an explanatory science of consciousness. This isn't a conclusion we come to lightly. But if our arguments are correct, then it is the prudent course. Many hope, and will continue to hope, for a satisfying explanation of consciousness. And these researchers will act on that hope – in vain. The only way to relieve this intolerable state (intolerable, at least, because it is a waste of time) is to give up wanting explanations of consciousness. That, we *can* control.

So, the problem of understanding consciousness is, for all intents and purposes, immortal. This explains why it is also ancient. Our modern problem of consciousness, of course, is not exactly the same problem as any of the problems ancient philosophers puzzled over, but theirs and ours share a deep family resemblance. For example, in his treatise, *On the Soul*, Aristotle wrestled with what is clearly recognizable as a version of the mind-body problem: he struggled to understand the nature of the soul and how the soul is related to the body. The scope of his notion of soul was wider than our contemporary notion of mind, and he apparently didn't distinguish between consciousness and cognition (but then, neither do some of our contemporaries), yet the problem with which he wrestled strongly overlaps with ours today.

There is another set of problems that is likewise ancient: the fundamental problems of philosophy – like the problem of freewill versus determinism, the

nature of the self, of moral responsibility, and the meaning of life, to name four. Like the problem of consciousness, little progress has been made on these problems. In fact, the morphological changes of these problems over the millennia, mirrors the transmutation of the soul-body problem into our modern problem of consciousness (e.g., McGinn 1993). So, perhaps all these philosophical problems share a common root or cause. Since consciousness, at least one's own consciousness, is so central to our existence (and is the most indubitable property of the universe), it is natural to wonder if somehow consciousness is involved in the production of all of the fundamental problems of philosophy. We will argue it is centrally involved.

6.1 The enduringness of philosophy: The proper view

Many have pointed out that philosophy's central questions never seem to get answered. Rather than getting solved, philosophy's problems are perennial: they may even have a recurring life cycle. In this, philosophy is seen to be very different from science and mathematics. Mostly, philosophers just ignore this aspect our their discipline. But once in a while a philosopher will wrestle with this conundrum. Colin McGinn is one such philosopher of recent vintage (Thomas Nagel is another recent one; we discuss him in the next section). McGinn has produced some keen observations as to philosophy problems' life cycle. But his explanation for why philosophy endures is substantially different from ours. McGinn views philosophy as being the result of human epistemic limitations (McGinn 1989, 1993). He writes:

> Philosophical perplexities arise in us because of definite inherent limitations on our epistemic faculties, not because philosophical questions concern entities or facts that are intrinsically problematic or peculiar or dubious. Philosophy is an attempt to get outside the constitutive structure of our minds. Reality is everywhere flatly natural, but because of our cognitive limits we are unable to make good on this general ontological principle. (McGinn 1993: 2)

This view makes philosophy our fault. Philosophy is the fallout of a failed intellectual endeavor; it aspires to be like science or mathematics, but can't. Smarter creatures, McGinn claims, might not be troubled by our philosophical problems – though they might have their own, which we, in turn, might find easy to solve (1989, 1993).

However, McGinn's view of the matter has missed a crucial distinction. McGinn thinks that philosophy is due either to some intrinsic property of

the problems of philosophy themselves, or to some limitation in us. He never considers the third option that philosophy's central problems are unsolvable because of a property of *all* intelligent, conscious cognizers, wherever and whenever they may exist in this or any other universe. If any intelligent being in the universe, anywhere, anytime, would be puzzled over philosophy problems, then it is misleading to construe philosophy as arising from human limitations. Rather, being puzzled about philosophy is an essential and deep epistemic fact about conscious cognizing. This third option is our view.

We think that philosophy isn't our fault; it's nobody's fault: philosophy is *not* a symptom of human limitation. *Any* conscious being or entity smart enough to wonder about its own consciousness is going to find the question of consciousness's supervenience enduringly puzzling. This is an epistemic cum metaphysical fact about knowers in our or any universe. And, because most of the other classical philosophical questions could be solved if we had a solution to the problem of consciousness (or vice versa), none of these classical questions can be solved either. (The "vice versa" is important. It means that in a certain sense to be explained below, the central problems of philosophy are equivalent.) We can put this another way: Most of the other classical problems of philosophy require for their solution dispelling their own kind of "Cartesian intuition", and since that intuition is ineluctable, no matter which kind obtains, none of the classical problems can be solved anywhere by any being or entity whatsoever. The existence of philosophy problems in the minds of other conscious, intelligent beings then, is, on our view, a necessary, epistemic fact about them, and a deep epistemological fact about the universe.

6.2 The Nagelian conjecture

Our view of consciousness' role in philosophy can be derived from Thomas Nagel's view. Nagel's great insight was that two fundamentally irreconcilable points of view – the subjective and the objective – are prominent in the generation of the enduring problems of philosophy (1979: Chapter 14, 1986). It is best to explain Nagel's view using an example. We will use the problem of freewill.

When our actions, say the writing of this book, are viewed from an objective, external point of view, they cease to be actions, and, instead, become events that occur. Hardcastle and Dietrich didn't write the book; they were caused to write it. They were caused to talk about it and sit at keyboards typing it. This book is the outcome of those events. As authors, we two can adopt this point of view ourselves: We didn't write this book; we were merely the last links in the causal chain leading up to it. When viewed externally, all human action (and indeed all agent-centered action) vanishes. Things don't get done; they happen. Nagel puts it well: "Any external view of an act as something that happens, with or without causal antecedents, seems to omit the doing of it," (1979: 199). However, the external viewpoint – the objective viewpoint – leaves something crucial out, namely, that an agent does it. I am writing these very sentences now. I am the source of them. They exist because of me. They didn't merely happen or happen through me. I created them. Again, Nagel: "[An agent's] actions appear to him different from other things that happen in the world, but not merely a different kind of happening, with different causes or none at all. They seem in some indescribable way not to *happen* at all..."(1979: 199, italics in original). Our actions don't just happen; we *do* them. We are the *sources* or *fonts* of action (1986).

There are, then, two vantage points from which to view these matters: agents, as fonts, accomplish things; there are no agents at all, but merely a universe of interlaced events, one after another. Nagel points out that, in the problem of freewill, neither point of view can gain the upper hand. The objective point of view, irredeemably leaves something out – agents, doers of deeds. And the subjective point of view seems incompatible with our robust, scientific view of our universe and us as material elements in it. The objective point of view is supposed to be truer, but how can that be if it leaves out agents?[1]

Nagel then argues that this phenomenon metastasizes. The self and its subjective point of view are *irreducible*, not just in the problem of freewill, but in the mind-body problem, the nature of the self and self-identity, in epistemology, metaphysics, and in ethics. The self and its subjective point of view are essential parts of the universe, and are here to stay. Nagel suggests that reality should not be identified with objective reality (1979: 211). Rather, reality is essentially *split* between the objective and the subjective. And unification is a vain dream. A final quote from Nagel sums up the situation: "[T]he consistent pursuit of greater objectivity runs into trouble, and gives rise to the philosophical problems [listed above], when it is turned back on the self, as it must be to pursue its comprehensive ambitions" (1979: 210).

We will assume that Nagel is right that the fundamental problems of philosophy all have the same underlying structure: a clash between irreconcilable points of view – the subjective and the objective views. In his 1986 book, Nagel clearly identifies this clash between viewpoints as the source philosophy (p. 6ff.). This is the claim we are interested in. Philosophy is the result of adopting first one and then the other point of view when asking some basic question (e.g., "Am I free?"). We enshrine this claim as the Nagelian conjecture:

The Nagelian conjecture: Philosophy is the result of switching between subjective and objective points of view when asking fundamental questions. These two points of view are irreconcilable and basic.

6.3 Deeper aspects of Nagel's conjecture

Nagel's insight must be pushed deeper if it is to function as an explanation of the cause of philosophy's central problems, and if we are to tie being conscious to being puzzled about philosophy's questions. We must ask: What is it about changing points of view that does the trick of generating philosophy? That objective points of view replace subjective ones and vice versa requires two very interesting mental processes as well as consciousness. Once we unpack this, we will be in a position to see how consciousness is related to philosophy.

The first process that we need to consider is what we call *referent maintenance*. As one's point of view changes, it is crucial that the referent of the point of view remain the same; otherwise, the subject will be unable to know that she is changing points of view at all. Consider walking around a table. As your points of view of the table change, it is crucial that you believe that you are always seeing the same table – that your beliefs about the table always refer to the same table. Any lack of certainty here would cause you to question whether you are having different points of view of the same object or views of different objects. It is crucial to the very existence of points of view that these two possibilities be distinguished and decided in favor of the former. So, if an author, puzzling over freewill, is viewing her writing of a book from her own, agent-centered, subjective perspective, and then considers her writing from an objective perspective, it is crucial that she believe that she is viewing from different perspectives – different points of view – the very same authoring of her book. In fact, this is a necessary condition on having points of view.

(1) If a subject believes she has points of view, then she believes she maintains referents across those points of view.[2]

(2) If a subject has points of view then she maintains referents across those points of view.

Secondly, there is the self. A point of view change doesn't make sense unless there is a single entity or subject whose viewpoint is changing. If you and I are standing on opposite sides of a table, you have your point of view and I have mine. There is no point of view *change*. There are merely two different points of view. To get viewpoint change, a single subject's viewpoint must first

be from one perspective and then from another. These two perspectives have to be occupied by the same subject at different times. This is why referent maintenance is required in the first place. This is another necessary condition on having changing points of view.

(3) If points of view change then there has to be a single subject having those points of view over time.[3]

Finally, we need to fold in consciousness. Consciousness, or conscious experience, plays a crucial role in grounding our *knowledge* of our beliefs about our points of view and the changes to them. We don't merely have the points of view, we have beliefs about them. When one is contemplating a philosophical problem – concentrating on how it is generated – one has beliefs about one's experiences from different viewpoints. For example, in considering that I am writing this sentence, I have beliefs about my own agenthood in causing this sentence to exist. When I change viewpoints and adopt an objective perspective, I have beliefs that I am merely a link in a causal chain. These beliefs are beliefs about my experiences. Hence they are warranted (at least in part) by the fact that I am having the relevant experiences themselves.[4] So, when one is contemplating the generation of a philosophical problem, one is conscious of the information from the relevant point of view, and the change in information as one changes one's point of view. And it is just this conscious experience that warrants the relevant belief. (However, one needn't be conscious that it is the change in points of view that is causing the problem. This just means that when one reads, e.g., Nagel 1979, one can be *surprised* and then *convinced* by his arguments.) We sum this up by saying, when generating a philosophical problem, you experience a point of view and the point of view shift, but not necessarily the shift in points of views.[5]

Nagel considers consciousness to be one of the philosophical problems his conjecture is designed to explicate. But it is crucial to see that consciousness itself plays an essential role in *generating* the relevant viewpoint knowledge in the first place.

These three ingredients – referent maintenance, a self or subject, and consciousness – need to be added to flesh out the Nagelian conjecture properly. We think that we also need to be explicit about insuring that the relevant beings or subjects have human-level intelligence as well as curiosity, since the absence of both these properties can be imagined in a cognizer of some sort. (Having human-level intelligence might entail having curiosity. But no one knows. So we add it explicitly.) Call this deeper, expanded version of viewpoint change *enhanced viewpoint change*. Then the improved Nagelian conjecture is this:

> *The improved Nagelian conjecture*: Philosophy is the result of enhanced viewpoint change when asking certain questions where the relevant viewpoints are the subjective and objective viewpoints (which are irreconcilable and basic).

All we have really done is elaborate, via logical implication, Nagel's original conjecture. If the arguments for his original conjecture work, then they work for the improved version. We think that the improved Nagelian conjecture is true.

Given our explication of the improved Nagelian conjecture, it should be obvious how consciousness helps create philosophy: It grounds our knowledge (our beliefs, if you prefer), about what we are perceiving from the relevant subjective and increasingly wide objective points of view.

According to the improved Nagelian conjecture (as well as the original) all the central problems of philosophy have the same structure. Hence, if we could figure out how it is that we manage to change points of view and what such changes really involve, we would unravel all these central problems. This suggests that the central problems of philosophy might stand or fall together (modulo the details about each problem, which, given a solution to their central, root problem, might not be hard to resolve). But in order to understand viewpoint change, we'd have to explain how consciousness is involved in viewpoint change in the first place (recall that that was one of our elaborations). And in order to do that, we very likely would have to reduce consciousness to something psychological, since viewpoint change is a psychological phenomenon. But of course, that's not in the cards. We can consider consciousness *sui generis* and just add it into our theory of viewpoint change, but this, as we have repeatedly mentioned, won't be very satisfying. Hence, it seems likely that viewpoint change of the kind in the Nagelian conjecture will remain unexplicated. This suggests that the central problems of philosophy will continue to stand – forever.

6.4 The nature and future of philosophy

It remains to say what philosophy is, on our view. As we just said, it seems to be a consequence of the improved Nagelian conjecture that the central problems of philosophy cannot be solved (Nagel is explicit about this claim). In a deep way, Nagel suggests, this is a good thing. He says:

> Certain forms of perplexity – for example, about freedom, knowledge, and the meaning of life – seem to me to embody more insight that any of the supposed solutions to those problems. (1986: 4)

We don't know if enduring perplexity is a good thing. But it does seem to be a fact about our existence.

So, the view we are promoting implies that philosophy does not make progress: if its problems have no solution, there is no sense to the notion of getting closer to a solution, which is one kind of progress. Indeed, there is no direction at all within the realm of philosophy: there are merely perennial puzzles. Some have criticized our view on this point by saying that we have rendered philosophy otiose. If this is the truth about philosophy (a view that would not surprise many lay-people), we should boldly embrace it.[6] We don't know if philosophy is otiose or not. We end this book with a discussion of how it may not be.

Making progress is just one way to judge a long-term human activity. It is certainly not the only way, nor even the best way. Consider art. Art makes no progress, arguably. The engravings in the cave of Combarelles I outside Les Eyzies de Tayac, France, made thirteen thousand years ago by people of the late Ice Age, are meticulous and delicately executed, and are every bit as compelling as something by Miro or Tanquy, Warhol or Dali. Of course, in many ways, the depictions of reindeer, bison, mammoths, and other mammals, along with geometric designs, don't speak to us as well as silk-screens of Campbell's tomato soup cans, or languid, dripping clocks. But the Ice Age people of ancient France were likely gripped by their art and the need to make it – as gripped as we are by ours.

This is obvious from examining the cave itself. Thirteen thousand years ago, the cave was extremely difficult to wiggle into, its opening forcing the artists to squirm along on their stomachs in the dank darkness for one hundred fifty yards, there to work for hours by the light of burning animal fat (deduced from the amount of soot on the cave walls). Their art spoke to them, we presume; our art speaks to us. We appreciate theirs; and they, if they could see it, might appreciate some of ours (although, the tomato soup can would probably leave them cold). But their art isn't *worse* than ours; our art isn't *better* than theirs. Their art is just different.[7]

Compare their art to their medicine. No one today wants to be treated by the shamanic techniques of thirteen thousand years ago – at least not exclusively, and not for anything serious. Modern medicine is a triumph of modern science, which does indeed progress. We know so much more medicine (and so

much in a different way) than Ice Age people, that it would be difficult to talk to them about their health, their world, and their lives, and equally difficult to express ours to them.

Art doesn't make progress because art is not solving a problem; art is not answering a question. We do not want to get embroiled in the philosophy of aesthetics – yet another unsolvable problem in philosophy – but it does seem as though art is not about knowledge so much as it is about expression (though of course one has to know something, in some sense, before one can create a work of art).

Certainly, the techniques for creating art of all kinds have progressed. Technology and skill, like science, do advance. Ice Age musicians would have no idea what to do with a synclavier (though given one, they might certainly produce something interesting). And it is true that techniques for creating such important artistic features such as a three-dimensional visual perspective emerged only a little under a thousand years ago. But even though the religious painters of the Middle Ages couldn't paint perspective well, it does seem as if their art did for them exactly what our art does for us, and it did it exactly as well.

It is reasonably clear, then, that not every important human activity makes progress, hence not every important human activity should be judged by how far along the path to ultimate answers it is. (Sports are perhaps another example. Though our fastest runners today are no doubt faster than the fastest Ice Age runners, the thrill of sport hasn't changed, and sports are arguably about the thrill.)

It might be, then, that philosophy is like art. Philosophy is *not* art, but it is like art in that it doesn't make progress, but is nevertheless a worthwhile, indeed essential, human endeavor.

So if philosophy doesn't make progress towards solving its problems, what is it doing? Clearly, philosophers *do* look for solutions to the great problems of philosophy. Many of them think that they *are* making progress. If they aren't, then what are they doing?

Philosophers are doing many things – at once. Philosophers recast the great problems for their time, unearth new avenues of inquiry, and propose solutions. This much is obvious. But the solutions always fail – some later philosopher always comes along and refutes earlier theories. We conclude that it's not the unattainable solutions that matter. What matters are the recastings and the unearthed avenues. These integrate new methods and discoveries with our sense of being human, our definitions of who we are. Philosophy's worth lies in the relation between its proffered solutions, the new methods and techniques of a time that the solutions tap into, and our views of ourselves. Philosophy is crucial to updating the definition of being human. Not producing, by increments, the "correct" definition, but updating it.

We can't just rely on chemistry, biology, psychology, anthropology, and the like to tell us what it is like to be human because these sciences leave out individual experience. Even the science of consciousness will do that. Sciences tell us what a human being is. Philosophy can tell us what being *us* is. Philosophy, then, is part of an ongoing definitional dialogue humans have with themselves both at a given time and across time.

For example, the indeterminism of quantum mechanics altered the debate about freewill. So did research in neuroscience. And so did the discovery of chaos and nonlinear, dynamical systems. By wrestling with the problem of freewill versus determinism using quantum mechanics, neuroscience, or chaos theory, we learned more about freewill and about the three new tools we were using to attack the problem afresh, and about us.

So, one answer to why philosophy might not be otiose is that it helps us understand what it means to be human as we acquire more scientific (and technical) knowledge. Viewed this way, philosophy is a sort of struggle to integrate human experience with increasingly robust and nonhuman (or transhuman) knowledge.

But all of this might sound either empty or highfalutin or both. Philosophy is part of a continual dialogue we all have with ourselves. So what? So is bathroom humor. What makes philosophy noble but bathroom humor not? We don't know. We are not even sure philosophy is noble. Are we learning anything by doing philosophy? Perhaps. We might be learning something by *doing* philosophy that is independent of actually solving its central problems. Science does progress (as does its associated technology). Though the problems of philosophy don't get solved, each time a philosopher offers a new solution to some grand problem we do seem to learn something, something about us and about our world.

But if we are learning about ourselves and our world by doing philosophy, why aren't we making progress, since we (and our universe) are finite? There is, it seems, only so much to know about humans. Hence, every time we learn something new, we get closer to the final goal of understanding ourselves completely. Once we know ourselves completely, won't we be able to solve the problems of philosophy?

This question confuses two kinds of knowledge. We can know all that it is possible to know about humans, without having a robust, unifying, *complete* theory that makes being a human across all circumstances totally understandable. That is, we can know all that it is possible to know about human beings

without knowing all there is to know about human beings. This is just to say that necessary epistemic limitations occur in understanding ourselves just as they do in understanding, e.g., the space of computable functions or the exact position and momentum of a fundamental particle. It is this latter kind of knowledge that will elude us. If our conclusions in the previous section are correct, then philosophy's problems are unsolvable, and, indeed, the problems themselves are necessary epistemological facts about our universe – a universe with intelligent humans in it. If you make the assumption that solving the problems of philosophy is required to finally and completely understand being human, then it follows that we will never finally and completely understand being human. So, no matter how many facts about humans we know, a deeply satisfying theory of being human in our universe will always elude us.

We think that this leaves us with a crucial truth. What matters now is the recognition that very important, deep questions *do not have answers*.[8] Besides the true statements and the false statements, we have to add the *undecidable* statements. Statements for which we desperately want answers, but for which we cannot get answers – and we can't get the answers for principled reasons. In mathematics, this situation is well known. There are crucial but undecidable propositions at the heart of set theory. And just as we get several kinds of set theory depending on, e.g., whether we assume the continuum hypothesis is true or not, so we get different kinds of philosophy depending on whether we assume, e.g., there is freewill or not.

What follows from this view of philosophy? We aren't sure. Certainly a kind of relativism seems to loom on the horizon. It might be that our conclusions render business as usual in philosophy otiose. It could well be that what we should do is fully adopt the view that the fundamental problems of philosophy are undecidable, and that there are many, many coherent philosophies. And then set off afresh, from there.

Appendix

Problems with zombies

A discussion of Chalmers's argument for dualism

1. Chalmers's zombies[1]

Chalmers's arguments against materialism and for dualism are unusual in that he does not rely on the notion of identity: psychophysical identity is never used. Instead, Chalmers relies on the notion of *supervenience*. Chalmers argues that consciousness only supervenes *naturally* on the physical: there is no way for consciousness not to arise in our universe, given our physical laws and our universe's initial conditions. However, he denies that consciousness supervenes *logically* on the physical properties of our world. He asserts that he can imagine worlds in which there is the same physical stuff as this universe, but no consciousness. These are the *zombie worlds* that philosophers are so fond of discussing. The key notion of supervenience for Chalmers is logical supervenience, since it is the failure of consciousness to supervene logically that makes dualism true, according to Chalmers.

This is the crucial part of Chalmers's argument against materialism: we can, according to him, coherently imagine a physical universe exactly like ours peopled with lively, bouncy but completely unconscious zombies, but we cannot coherently imagine a physical universe exactly like ours at the fundamental particle level without also imagining molecules, insects, penguins, and governments. Because of this difference in what we can coherently imagine (or conceive), it must be that consciousness is different from atoms, molecules, etc. In fact, it must be that consciousness is not really physical after all. Consciousness, say the dualists, seems to be a further fact over and above the physical facts of our universe. This is Chalmers's position.[2]

Here now, in short form, is his general argument against materialism.

1. In our world, there are conscious experiences.
2. There is a logically possible world physically identical to ours in which the positive facts about consciousness in our world do not hold.

3. Therefore, facts about consciousness are further facts about our world over and above the physical facts.
4. Therefore, materialism is false (Chalmers, p. 123).

Perhaps it is now intuitively plausible that consciousness doesn't logically supervene. Given that all the positive facts about the world do logically supervene, consciousness emerges as quite peculiar indeed. If all this is correct, then we are required to reassess our metaphysical assumptions about the nature of the world.

We discuss two common objections. Many people suggest that consciousness' not logically supervening is no cause for alarm, and certainly no cause for the drastic measure of accepting dualism, because *nothing* logically supervenes on the physical. Researchers make this objection because they claim to be able to imagine a world very different from ours somehow emerging out of all the low-level facts of our actual world. The other objection is that they genuinely cannot imagine a zombie world and so, in truth, there are no such worlds (we've given that objection ourselves at times). They agree that consciousness has yet to be explained, but, when it is, we will see that it is reductively explainable and not surprising at all, given the low-level facts.

Chalmers's response to both these groups is to urge them to be more careful about using their imaginations (or, using their concepts, as he prefers to put it). Playing with possible worlds is not a game for amateurs. Presented with these two objections, Chalmers just goes through the development of the intuitions carefully, trying to convince the skeptics of the error of their ways.

However, note that the presence of these two *contrary* intuitions means that there is more to imagination and intuitions about what's possible than meets the eye. Both of these objections can't be true (but they can both be false). The fact that we have heard both advocated by intelligent and well-informed researchers, as well as Chalmers's intuition that both objections are, in fact, false, inclines us to think that intuitions about what is logically possible are very delicate, if not outright suspect. We will say more about this below.

Still, in practice, it is usually possible to make plausible to the first objectors that virtually everything *but* consciousness logically supervenes, and make plausible to the latter objectors that consciousness *alone* doesn't logically supervene.

2. The Kripkean view of worlds

Possible worlds are crucial to Chalmers's arguments against materialism and for naturalistic dualism. Possible worlds are part of a large and venerable project in philosophy called *modal metaphysics*, a project whose goal is to develop theories of necessity and possibility (or contingency), specifically, to develop theories about which things are necessary and which are possible, but not necessary.

Transworld identity is a crucial issue in modal metaphysics: What is the criterion for identifying individuals across possible worlds? It is our view that the answer one gives to the problem of transworld identification is a function of how one settles the general ontological question about possible worlds. For instance, Jaako Hintikka does not appear to be bothered by the problem at all, because he takes 'possible worlds' terminology metaphorically. Yet, Hintikka has realized to his "considerable consternation that the likes of David Lewis and Alvin Plantinga [both of whom do see a problem to be solved] are taking the metaphor literally" (1993).

In general, researchers can take their modal commitments one of two ways: (1) realistically, in which case modal talk refers to existent possibilia; or (2) nominally, in which case modal talk does not refer to the existence of ways things might have been but rather represents turns of phrase (i.e., a metaphor to help cash out a particular formal apparatus). Hintikka appears to be in the latter group. Chalmers, however, is some sort of modal realist. In particular, he says that our formalisms for capturing modalities are logically dependent upon our modal-metaphysical intuitions.

There are, of course, varieties of modal realism. If we exclude Kripke's view for the moment (1972, 1980), realists minimally contend that sentences are necessary or contingent depending on whether a sentence is true at all possible worlds or true at some world or other, respectively. This, on the realist view, in turn commits us to admitting independent entities (viz., worlds) into our ontology. It is the independent existence of such entities that allows our modal talk to be genuinely referential. Lewis's variety of realism, dubbed "Lewisian mad-dog realism" by John O'Leary-Hawthorne (1996: 198), holds that possible worlds are concrete physical counter-parts to our own but that they exist in an alternative space-time dimension (1986). This is clearly the extreme case, but the point to be made holds for the entire class of modal realists: As soon as we admit worlds as independent entities, then we must face the epistemological issues of that conviction, namely, how do we know what these possible worlds are like? In particular, how can we tell who's who in a possible world?

More technically, once we commit ourselves to modal realism, the problem of transworld identity arises. And the proponent of modal realism owes us some solution or other.

Chalmers does not want to talk about identification across possible worlds. Thus he writes that, in general, he

> will not be concerned with questions about whether individuals in those [other possible] worlds might have different "identities"... These issues of transworld identity raise many interesting issues, but are largely irrelevant to my uses of the possible-worlds framework. (1996:367, Note 30)

Chalmers does not think he owes us a solution to the problem of transworld identity, and he does not think the problem of transworld identity is relevant to his enterprise. If he does not find this problem relevant to his project, then (since modal realism is what makes the problem meaningful) we have that Chalmers is not a typical modal realist about possible worlds. This is more than mere infelicity: we know that Chalmers wants to be a realist of some sort with respect to our modal talk, since mappings are obviously real for him (they have to be, so that he can define identities between worlds), and mappings obtain between individuals in worlds instantiating certain properties; but on the other hand, given his reluctance to address transworld identity, we conclude that he is not a typical realist about possible worlds.

Now Chalmers needs a way out. Is there a sort of realism that avoids the problem of transworld identity? Kripke, it seems to us, is just this sort of realist.

That Kripke is some sort of realist when it comes to understanding our modal discourse is easily verified. In explaining how it is he understands "worlds talk", Kripke writes:

> But I do not wish to leave any exaggerated impression that I repudiate possible worlds altogether, or even that I regard them as a mere formal device. My own use of them should have been extensive enough to preclude any misunderstandings. (1980:16)

Given that possible worlds are more than mere formalism for Kripke, it is clear that he is not a modal nominalist. Kripke's brand of realism, to be sure, is the most modest strain. He makes the distinction between his interpretation and other modal realists in terms of an analogy to elementary probability. Given two distinguishable dice A and B, each fair and six-sided, the probability of throwing an eleven is easily computed. Eleven could turn up in exactly two ways: die A is 6 and B is 5, or A is 5 and die B is 6. Since each die has six sides, $P(11) = 2/36 = 1/18$. In order to figure this and similar probabilities, we discuss

possible outcomes quite routinely. Kripke's point is that we can do this, and do it meaningfully, without positing "that there are some thirty-five other entities, existent in some never-never land, corresponding to the physical object before me" (1980:17).

For Kripkean realism, there is no problem of transworld identity to solve. If we are considering some counterfactual situation about, say, John Nash – perhaps one in which he never studied mathematics and never won the Nobel Prize – there is no transworld identity problem. Witness: there is no issue of us traveling to a distant galaxy in which someone resembling John Nash in certain ways never studied mathematics, forcing us to ask, "How do we know that *this* is John Nash?" Rather, by virtue of the fact that the counterfactual situation under consideration is one about *John Nash*, the problem just dissipates. This is precisely what Kripke means when he writes that "'possible worlds' are stipulated, not discovered by powerful telescopes" (1980:44, for a good discussion of the nature of this stipulation, see Salmon 1996). So, following the Kripkean line: there is no problem in stipulating that a discussion regarding how things might have been for Nash in some contrary-to-fact scenario is still a discussion about Nash in some contrary-to-fact scenario. This is called, by Kripke, *rigid designation*, and the name *John Nash*, is a *rigid designator*.

The Kripkean view of possibilia seems to fit Chalmers rather well. This is a way for Chalmers to embrace realism but eschew transworld identification altogether. The match is not quite as felicitous as it seems at first blush, however. The difficulty lies in the fact that Chalmers wants to describe worlds purely qualitatively, and Kripke thinks this is an unnecessary restriction. Recall Chalmers's definition for logical supervenience. For the argument against reductionism using the logical possibility of zombie twins, he wants to qualitatively specify a world identical to our own with respect to all microphysical facts, but differing with respect to phenomenal facts. In the characterization of supervenience that Chalmers gives, it is not clear that there are any rigid designators at all, and rigid designation is what is doing the work on the Kripkean view. This seems like an insuperable difficulty between the two positions, but in fact it is not.

For supervenience in general, it may well be the case that qualitative descriptions and their identities are what are in order. But, we think, the case for consciousness as Chalmers sees it is different. Chalmers takes conscious experience to be "at the very center of our epistemic universe" (1996:74). It is from this center that his zombie thought-experiment originates. Since each of us can only know (directly, incorrigibly, etc.) about our own private phenomenal experiences, Chalmers cannot ask us to imagine a world physically

identical to ours (in the qualitative sense, of course) but in which everyone else lacks consciousness. Modulo our own private experiences, that could be this world for all we know.[3] Rather, he is asking each of us to imagine our own zombie *twin*: someone qualitatively identical to ourselves, but who lacks phenomenal experience (see, Chalmers 1996: 95ff.). His anti-reductionist argument is not "imagine all the same B-facts, without the A-facts holding", as he in places mistakenly poses it, but instead "imagine a possible world inhabited by your zombie twin, identical to you with respect to physical facts but a zombie nonetheless." We can easily imagine such a world inhabited only by our "twins" with varying degrees of similarity to ourselves without having phenomenal experiences. Suppose your zombie twin had a world-contemporaneous *non*-zombie twin. Which "twin" would Chalmers consider as counting for his thought experiment? Chalmers is likely to reply that, by stipulation, the one without phenomenal states is the one that matters. And this is precisely the Kripkean line.

So, when we are imagining zombie twins, we are simply imagining counterfactual situations about ourselves in which we do not have qualitative phenomenal experiences. And, taking names and demonstratives to designate rigidly, we then may conclude that the Kripkean view of stipulation is indeed correct for analyzing the Chalmerian supervenience relation in the case of consciousness.

So now we have the most charitable read of Chalmers's position as relying on Kripke's idea of stipulation, and thus at bottom his argument for the non-supervenience of consciousness does not in fact rely on qualitative description of worlds. Chalmers can resist this interpretation of his position only if he provides us with an alternative, and that would require providing us a solution to the transworld identity problem.

3. Consciousness and conceptual truth

Now we are ready for the next difficulty. For Chalmers, the formalisms for capturing modalities are logically dependent upon our modal-metaphysical intuitions, and these, in turn, depend on what is *conceivable*. So, the step from the conceivability of zombie twins to their logical possibility is crucial for Chalmers. If he cannot tie these two together, then his thought experiments (e.g., imagining a zombie twin world) are of no consequence to reductionistic theories of mind.

The traditional theory of meaning and reference, handed down since Frege, holds that each concept, c, determines a function f_c: $W \rightarrow R$, from the set of all worlds to the set of all referents, such that when applied to some world, $w \in W$, $f_c(w)$ yields an *extension* of the concept. The function itself is known as the *intension* of the concept. Chalmers wants to use Kripke's insight here:

> No single intension can do all the work that a meaning needs to do. The picture developed by Kripke complicates things [for the traditional view] by noting that reference in the actual world and in counterfactual possible worlds is determined by quite different mechanisms. (1996: 56–57)

From here, Chalmers splits the class of intensions into disjoint subclasses: primary intensions and secondary intensions, which correspond to the two ways in which reference is fixed. For the primary intension, we have reference as it is fixed in the actual world, and for the secondary intension, we have reference as it is fixed in counterfactual situations, given that reference is already fixed in the actual world. So the primary intension of "water" is "that clear, potable stuff, coming out of taps and found in rivers and streams, etc." (or something similar, Chalmers uses "watery stuff" 1996: 58ff.). The secondary intension typically comes from science, for it depends on how things are in the actual world. We investigate that clear potable stuff in this world and discover, after some experimentation, that all analyzed instances of it coincide with instances of the structure H_2O. So we conclude that water is H_2O. The secondary intension, then, of water is H_2O.

Focusing on the primary intension, water (that clear, potable stuff) did not have to be H_2O. That clear potable stuff could have been some other chemical, XYZ, instead of H_2O, had the universe been structured differently (cf., Putnam 1975). On the other hand, given that the clear, potable stuff in the actual world's lakes, oceans, and rivers, etc. *is* H_2O, "water" picks out H_2O in all possible worlds; there is no possible where water (H_2O) is XYZ – that is incoherent (this is the Kripkean line). Here, it is the secondary intension of "water" doing the referential work. Note, neither the primary intension, nor the secondary intension is *the* meaning of a term (say, "water"). Which is the meaning depends on context and what one is trying to accomplish.

Secondary intensions are intimately related to Kripke's necessary *a posteriori* truths, and primary intensions cohere in the same way to *a priori* necessity. Chalmers says: "Kripkean *a posteriori* necessity arises just when the secondary intensions in a statement back a necessary proposition, but the primary intensions do not" (1996: 64). So, since "water" could have picked out XYZ had the actual world been different, "water is H_2O" is not necessary using the primary

intension of "water", but given that the actual world turned out the way it did "water is H_2O" is necessary using the secondary intension of "water." "Water is that clear, potable stuff" is necessary when evaluated from the primary intension. As Chalmers puts it, this is a "conceptual" connection. We can't know, *a priori*, that water is H_2O, but we can know, *a priori*, that water is that clear, potable stuff. These identities hold necessarily, for they turn on how our language and how our concepts work, not on how the world is. As long as we speak the same language, then the primary intension of "water" will be the same and we can know the primary intension without investigating the world (much). All we have to do is know the concepts (learning which, of course, requires investigating the world, at least for humans).

So too for the concept of consciousness. Our primary intension of "consciousness" picks out, refers to, our subjective experiences. Note, that for "consciousness" (and maybe only for "consciousness") our primary and secondary intensions coincide (1996: 133). In the actual world, consciousness just is phenomenological experience, subjective experience. So, the primary intension of "consciousness" is "phenomenological experience" (or "conscious experience"). (We might say that the primary intension for one's conscious experience is that it have a certain phenomenological feel.) However, this is just what the secondary intension picks out, too. Given that "consciousness" picks out conscious experience in the actual world, this what it picks out in all possible worlds. Contrast the case with water. For "water"'s primary intension, we use our *a priori* concept of water: watery stuff – that clear, potable stuff. But for the secondary intension, we use how the actual world turned out to pick out water in counterfactual worlds. In the actual world, water turned out to be H_2O. So the primary and secondary intensions pull apart. But this doesn't happen with consciousness. *A priori*, consciousness is conscious experience. But upon deeper analysis, this is what consciousness turns out to be, too. So in counterfactual worlds, consciousness is also conscious experience. Hence the secondary and primary intensions coincide.[4]

We note Chalmers's great faith in the *a priori*. His surety in the *a priori* stems largely from this distinction in intensions. And since his argument against reductive/functional explanations of the phenomenal is made on *a priori* grounds, he is most concerned with primary intensions. The first thing he must do is flesh out just how we are to make sense of, e.g., "watery stuff". This is crucial. The primary intension of "water" is going to be something very close to a definite description. The only codicil Chalmers makes is that we don't confuse the description that is fleshing out the intension with the function itself. This point about definition descriptions will be important in the next section.

With the apparatus of primary and second intensions in place, it is a short step to two notions of possibility: 1-possibility and 2-possibility, one corresponding to each type of intension, respectively. Conceivability, then, is also divided into two classes: 1-conceivable situations and 2-conceivable situations. Since he is interested rejecting materialism on *a priori* grounds, he needs to connect 1-conceivability to 1-possibility. A situation is 1-conceivable on Chalmers's view if it is conceivable according to the primary intensions of the terms involved. And, if something is 1-conceivable, it is 1-logically possible.[5]

Now, finally, we are ready to argue that zombies are not possible. This claim falls directly out of the Chalmerian picture of primary intensions, 1-conceivability, and 1-possibility. Here's how.

4. The impossibility of zombie twins

Remember that a possible world inhabited by your zombie twin is a counterfactual situation about *you* – one in which you lack phenomenal experiences. This seems readily conceivable to most people, especially philosophers. But this initial intuition is not strong enough for Chalmers. He needs it to be 1-conceivable, and thus 1-logically possible. So we must turn to the primary intensions of the concepts involved in a situation in which we are not conscious. In so doing, however, we find that our zombie twins are not 1-logically possible. Note that because Chalmers prizes the epistemic asymmetry of conscious phenomena, each of us will be describing a different (purportedly) 1-possible situation: Chalmers will be checking the 1-possibility of a situation in which Chalmers is not conscious, Dietrich will be checking one in which he is not conscious, and Hardcastle will be doing the same for herself, and so on. In each of these situations there is an indexical concept: "Chalmers" in the situation he is imagining, "Dietrich" and "Hardcastle" in their appropriate cases, etc.

From what we know about primary intensions it is unproblematic to say that the primary intension of "Chalmers" is "chalmersy stuff", that of "Dietrich" is "dietrichy stuff" and similarly for "Hardcastle". But what is chalmersy stuff, anyway? Saying "Chalmers is chalmersy stuff" does not provide enough information about the primary intension of this concept to evaluate the possibility of the situations at hand. In order to get the correct analysis we must turn to something like definite descriptions, taking care not to confuse the description with the function itself. Notice that such descriptions are only relevant to what the primary intension of "Chalmers" is *for Chalmers* (and likewise for Dietrich, Hardcastle, and so on), because what we want to know is whether the

subject can imagine a counterfactual situation about herself in which she does not have conscious experience, while her psychological life remains unchanged.

Now, it is reasonable that whatever the description is that Chalmers uses to flesh out the primary intension of "Chalmers" – i.e., to give substance to "chalmersy stuff" – it must have consciousness as a constituent member. Thus he writes that, "our core epistemic situation already includes conscious experience" (1996: 195). But then we have that Chalmers's conscious experience is part of the primary intension of "Chalmers", and the same for Dietrich and Hardcastle, etc., at least as evaluated by each of them. And if that is the case, then since all of this was to evaluate the logical possibility of counterfactual situations about ourselves, it follows that our zombie twins are not 1-possible – there is no possible world where you are a zombie. The primary intension for each of us – i.e., for each of our self-concepts – *includes consciousness*, and this, in turn, requires that we each be conscious in all possible worlds in which we exist. Hence, although zombie twins appear to be readily conceivable (in some sense), given Chalmers's analysis of primary intensions, they are in fact not 1-logically possible. So either the connection between conceivability and possibility that Chalmers crucially needs is not a tight one at all, or the connection is tight and (by modus tollens) zombie twins are not 1-conceivable (though we may think they are, in some other, possibly vague or loose sense of "conceivable").

Note that even this much undermines Chalmers's reliance on conceivability. For the first horn of the dilemma, conceivability simply does not imply possibility, even restricting the situation to 1-conceivability and 1-possibility. This means that, contra Chalmers, matters of what is or is not possible are *not* accessible from the armchair. For the second horn, conceivability does imply possibility, but we can't trust what we consider conceivable: though zombies seem 1-conceivable, they aren't, in fact. Matters of what is conceivable become very slippery. Again, armchair metaphysics is called into question.[6]

A quick way to summarize our position is to consider again the synopsis of Chalmers's argument presented in section 1 (and in Chalmers 1996: 123). Our claim is that the only case that is relevant to the truth of premises 1 and 2 is one's own case, and in that case, though you can know premise 1 is true, you also know that premise 2 is false, given Chalmers's technical modal machinery. For properly fleshed out, premise 2 requires the 1-possibility of zombie twins. And they aren't 1-possible (it is very important to adhere to the technical notions and technical vocabulary, here: the crucial notions are 1-possibility, 1-conceivability, and primary intensions). In the only case that

matters, one's own case, premise 1 is true but premise 2 is false. Hence, the argument is unsound.

Not only is the argument unsound, but the reasons for its unsoundness – the 1-impossbility of zombie twins – renders dualism likely false. Since we all are restricted to using primary intensions, any conscious creature or entity can only use itself in these modal deliberations. I am conscious. So, the primary intension of my self-concept includes my consciousness: I am *essentially* (1-essentially) conscious. This generalizes: any conscious creature or entity is restricted to using its own self as the relevant case, and from that personal, idiosyncratic perspective, it is essentially conscious. Hence, there are no zombie twins; they are 1-impossible. This is sufficient to render any arguments for dualism that rely on zombie intuitions false, for we know that at least consciousness exists, i.e., in any world where you exist, at least your consciousness exists; this is what our premise about consciousness being part of one's primary intension (one's self-concept) amounts to.

Three objections to our argument need to be handled. First, one might insist that we can in fact easily conceive (according to 1-conceivability) zombie twins of each other, and thus zombie twins are logically possible. A second objection is to concede the impossibility of zombie twins, but maintain the 1-possibility of zombies in general. Finally, one might object that the situation calls not for the primary intension of, e.g., "Chalmers" but for that of "physical Chalmers" – i.e., all the physical properties of Chalmers. None of these work.

That we can conceive of, e.g., Chalmers's zombie twin, and that he can conceive of ours, is of no consequence. The most salient way to make this point is to note that as far as each of us is concerned, we in fact *do* inhabit a world in which everyone is a zombie – everyone except oneself, that is. Changing the problem of consciousness to the problem of other minds will not save Chalmers's argument.

Nor will it do to concede the impossibility of each of us having a zombie twin, and try instead to make the case from the possibility of zombies in general. (For example, one might claim that all one needs to do is to imagine brain processes occurring without consciousness occurring.) There is exactly one case in which one can legitimately and non-question-beggingly imagine a zombie: one's own case. Outside of that one case, intuitions are useless. If forced to consider non-twin zombies, the materialist will simply retort: "Those zombies are just not physically enough like us to have conscious experience; they're not physically identical to us, and so of course it is possible that *they* are not conscious. But what does that prove?" For Chalmers to pursue this line of defense would require shifting the discussion away from physically identical

creatures to physically *similar* creatures, and then the whole notion of logical supervenience would need overhauling. You have to imagine your body and brain processes occurring without your consciousness. This you can readily do, no doubt; so can we; we freely admit it. But this as we argued is due, not to the actual logical possibility of zombie twins, but rather to some interesting facts about your (and our) concepts about consciousness and your (and our) abilities to conceive of possibilities in general. So, if Chalmers's argument can't work for zombie twins, then it can't work anywhere, for it loses all of its intuitive force.

Finally, objecting that what needs evaluating is not the primary intension of "Chalmers" but rather that of "the physical make-up of Chalmers" begs the question against us. Since Chalmers is imagining a counterfactual situation in which he has no phenomenal experiences, saying that he needs only to imagine the physical respects of Chalmers already assumes that he can truly separate the physical from the phenomenal.[7] Put in his technical vocabulary, Chalmers is likely to respond that all he need consider is the primary intension of his physical duplicate. But again, this is question-begging, for conceiving (in some sense, but not 1-conceiving) of one's mere physical duplicate already assumes that consciousness does not logically supervene on the physical.

But what if someone insisted that she could imagine her brain processes occurring without her consciousness occurring, and that this was sufficient to get the Chalmerian argument against materialism started? She can just "see" that her consciousness and her brain processes logically pull apart. This move is consistent with the restriction that Chalmers's anti-materialist argument use zombie twins. And one can set up this objection so that it is consistent with our conclusion that zombies twins are not possible. There is no possible world where *you* are a zombie (for, evaluated in terms of 1-conceivability and 1-possibility, you essentially have the property of consciousness). But perhaps there could be a world where your brain processes do not result in your consciousness. In this case, we are not imagining that we have zombie twins (which we do not have), but rather that our brain processes have zombie analogues. This move is still question-begging because it assumes the very separation between consciousness and brain processes that it tries to prove. Furthermore, we have argued that it simply isn't 1-conceivable that one's brain processes could occur without one's consciousness thereby occurring, and it isn't 1-possible that such a thing could occur.

We admit, as in the zombie twin case, that we can readily *imagine* our brain processes occurring without our consciousness occurring, but is this kind of imagining the same as 1-conceivability? Not according to our analysis above, but in all honesty, we don't really know. For one thing, we aren't sure 1-conceivability is a true psychological kind. The definition of 1-conceivability depends crucially on the notion of primary intension. The definition of primary intension depends on how reference is fixed. But there are no well-accepted theories of reference; cognitive scientists and philosophers aren't sure how reference works. One specific problem here is this: reference clearly involves concepts and categorization and there are no robust, well-accepted theories of concepts.

Another closely related problem is that there may be many kinds and grades of conceivability beyond just 1- and 2-conceivability. Again, without a theory of concepts and conceivability, it is hard to say, one way or the other. Though psychologists know quite a bit of experimental detail about concepts and how they are formed, a theory making sense of all of this detail is still a long way off. Therefore, putting so much trust in concepts and conceivability at this stage is premature. Lastly, it might be that our imaginations are conceiving the well-known epistemic gap between consciousness and the physical world of our brain processes and invalidly deriving a metaphysical gap from that.[8] So, while we aren't completely confident that dualism is false, at a minimum, we can say it certainly seems that we don't have zombie twins, and there seems to be no non-question-begging way of focusing on our bodies alone sans their consciousnesses.

We are back to the dilemma raised earlier: for the case of consciousness, what is conceivable (1-conceivable) is not a reliable guide to what is possible (1-possible); or zombie twins are not 1-conceivable contra intuition. We are impressed that picking between these two is difficult. This suggests that when it comes to consciousness, our modal intuitions should not be trusted. When intuitions are untrustworthy it is often because we cannot get clear on the necessary details. Hence, when we are imagining zombie twins we must be imagining vaguely somehow or in some other way.[9]

5. Conclusion

There is nothing obviously wrong with this argument. It certainly seems correct to us, at least in so far as we can bring ourselves to trust our modal intuitions. Yet, this argument does nothing to defang or even mitigate our Cartesian and zombie intuitions. These are as strong as ever in us. And in our reader's, too, no doubt.

Notes

Chapter 1

1. Of course you might misinterpret or mislabel certain complex conscious states, e.g., you might confuse infatuation for love or fear for hatred, but you are still experiencing. More simple experiences, like seeing color, are veridical: it is unlikely you aren't experiencing seeing red when you seem to be seeing red. There might be nothing red in your field of view, but that doesn't mean you aren't seeing red.

2. We say our catalog is good because we are able to achieve so much with it, such as building airplanes, the Sears Tower, the Internet, and treating leukemia and diabetes. In addition we are often able to see what sort of knowledge holes we have and why we have them. We can't cure AIDS, but we have a good idea why this is the case, and indeed what, in general, it would take to remedy this problem.

3. Disputes and debates in consciousness research frequently trip over the fact that researchers only have, at best, necessary conditions. Researchers often use the fact that proposed analyses are necessary but not sufficient to argue against their competitors' preferred approach without recognizing that they are following the same strategy (see, e.g., Taylor 1998, 2001). So, there's a positive methodological consequence to recognizing that consciousness researchers only have, at best, necessary conditions: namely, pointless disputes can now be avoided.

4. In fact, mention consciousness or the problem of consciousness to almost any cognitive psychologist and you will frequently get an abrupt change of topic.

5. This term originally came from Flanagan (1992), but has since moved into common usage. Even *Time* magazine has used the word.

6. It is somewhat hard to find avowed mysterians out there (though some postmodernists seem to fill the bill, but this might be because they regard *all* science as impossible). This is mostly due to the fact that the very distinction we want to draw is not being drawn. There are many who regard consciousness as forever beyond human understanding, as forever unexplainable; Colin McGinn and Jerry Fodor are just two famous examples (e.g., 1999, 1994). But it is not clear which subset of researchers go on to conclude that a science of consciousness is impossible. Arguably, at least some of them do (Fodor, for example). But whether there might be a nonexplanatory science has never been considered by them, so if given this out, some might opt for it. We will be arguing that this option is the best available. On the other hand, naturalists are everywhere. Churchland is a good, indeed, classic example(see, e.g., 1989). Churchland assumes that the only kind of good science is reductive science. So

he argues that there will be a future, new science of consciousness (and its fundamental conceptual changes) *and* it will allow us to deeply understand and explain how seeing red is such and such a brain state (perhaps described using some new and powerful neural vocabulary). He has missed the distinction we are after: a science needn't reductively, and certainly not satisfyingly, explain a phenomenon to theorize about it.

7. Dietrich has since modified his position.

Chapter 2

1. We recommend seeing the original *Matrix*. It's as good an intuition pump for the Cartesian intuition as there currently is.

2. Both intuitions also can be couched in terms of twins, zombie twins in one case and Cartesian twins in the other. Your zombie twin is you in a possible world where you have no phenomenal states (but see the Appendix). A Cartesian twin is you in a radically different possible world who nevertheless has your same experiences. For example, in that other world, the sky really is yellow (the actual you would see it as yellow), but your twin sees it as blue – the very blue you see in this world. Your Cartesian twin is your twin in virtue of having your experiences. Your zombie twin is your twin in virtue of being made of the same physical stuff as you.

3. Two separate processes result in mountain building: folding and faulting. But geological, reductive explanations require distinguishing between the two as types. Bat wings and bird wings both enable flight. Here reductive explanation seems to be handled by one thing: lift. But bats fly differently than birds. If one wants to explain why bats flutter so much, one will have to invoke the detailed differences between bat wings and bird wings as types. Chalmers discusses this issue (1996:48). We disagree with him that supervenience can be made sufficient for reductive explanation by distinguishing between illuminating explanations and mystery-removing ones.

4. The argument in the appendix against zombie twins does not show that dualism is false – it merely shows is that there are no zombie twins. To get from this to the falsity of dualism, one would need this premise: "If dualism is true, then zombie twins are logically possible." But this premise could be false. For example, the putative existence of zombie twins is supposed to show that consciousness doesn't logically supervene on the physical. But there could be different epistemic kinds of logical supervenience. *Kind one* is accessed conceptually, and *kind two* is accessed only evidentially or abductively. The first kind is the typical kind: most philosophers who use the tool of logical supervenience assume that to say that *A* logically supervenes on *B* is to say that there is conceptual relation between *B* and *A*. Also typically, such philosophers see a tight connection between the possible and the conceivable. But *A* could logically supervene on *B* even though we can't conceive of how that could be the case. We might be led to conclude such supervenience exists because it is the best explanation available. In this case, there would no tight connection between the possible and the conceivable. Kind two is compatible with a very robust sort of dualism. Consciousness might really be a nonphysical property that nevertheless logically supervenes on the physical. Oddly enough, Chalmers himself shows how this might be true in Chapter 8 of his (1996) book. There he develops a sort of dual-aspect theory using information as the fundamental constituent of the universe, where information has both a physical and a phenomenal aspect. In this case, it is quite natural (or at least, possible) to see the phenomenal aspect as logically tied to the physical aspect: dualism is still true, but there are no zombie twins. Chalmers would probably insist on calling kind two *natural supervenience* (1996:36). He would say that there is no reason to conclude that logical supervenience is at work if one cannot cash out the supervenience relation conceptually. But this assumes a tight relation between the possible and the conceptual. One can reasonably deny that: one can, for example, insist that the Cartesian and zombie intuitions are persistent illusions. Doing this opens up the possibility that zombie twins are impossible but nevertheless conceivable – perhaps even necessarily conceivable (any conscious creature of sufficient intelligence is going to have the intuition that zombies are possible, even though they aren't). Finally, our argument doesn't show that dualism is false because given that dualists aren't convinced by it, all it might really show is that one's modal intuitions are easily muddled. In that case, arguments for or against zombie twins are not to be trusted. We think this possibility is very likely: see the Appendix.

5. We are assuming for simplicity that there is a unitary neural correlate of consciousness. But there may not be a single NCC, there may be many, because there may be many consciousnesses in a single human (see, e.g., S. Zeki 2003). If this were correct, however, this wouldn't affect our argument since, for our purposes, we could draft all the relevant NCCs into one large NCC.

6. We deliberately use the verb "see" here because the best way to relegate the zombie and Cartesian intuitions to persistent illusion status is to counteract them with more compelling *perceptual* information.

7. Some cognitive scientists who see all of perception as involving at least some inference claim that this objection actually shows that we can see the supervenience relation directly, in the only sense possible, for on their view, perception is never really "direct", but always proceeds via inference: "direct" then means something like "immediate inference." But there is clearly an important distinction between, e.g., seeing a car move and seeing its drive train rotate and concluding that the latter causes the former: the seeings are different from the inferring. So we will continue to draw the distinction between directly seeing and inferring.

Chapter 3

1. For a dissenting opinion on the issue of logical inference and conceptual analysis in reductive explanation, especially as it applies to consciousness, see Block and Stalnaker (1999). Chalmers and Jackson (2001), is a response to the Block and Stalnaker paper, but Chalmers and Jackson focus mainly on the issue of whether or not the relevant logical inferences are *a priori*. Chalmers's and Jackson's view is similar to ours: reductive explanation does require logical entailment. We should mention that there are scientists who are confused about this matter, especially in psychology. It is not uncommon to hear psychologists – even compu-

tationalists – say that fixing the low level computational states of brains is not sufficient to fix the high-level cognitive states. Sometimes this is because they confuse the notion of implementation with inference. But most often, it is because psychologists misunderstand the nature of logical supervenience and the inferential role of concepts in it.

2. Fodor, the Don Quixote of cognitive science, has argued against this extremely plausible orthodoxy; he thinks concepts are *not* used for categorization and recognition. See his (1998a: Chapter 4; and 1998b). For a response, see Dietrich (2001) and Giesy and Dietrich (2001).

3. The metaphoricalness of these phrases is probably an artifact of the imprecise language we have to use to describe what is going on in our heads as we adopt various points of view. The reason our language is imprecise in this area is that its perceptual component and its structure were designed for external use, primarily. We see a mountain from its eastern flank. We hike around to the north and see the same mountain from that direction, gazing at its north face. These are paradigmatic cases of changing points of view: we literally move the perspective from which we view the mountain by moving our body and hence our eyes. Human languages have mechanisms that are very good at describing cases like these. If, as seems likely, we pressed such mechanisms and such language use into service for describing our mental "perambulations" then it is no accident that terms for describing an external world are found metaphorically describing an internal one. For some interesting theoretical discussion on this, see Deacon (1997). McGinn (1993), also has a nice analysis of this phenomenon.

4. On Nagel's view, it is possible that one subjective point of view could be more objective than another. On our view this is impossible. This suggests another argument for our view of points of view: it is ontologically less profligate than Nagel's. It is cleaner to posit two completely distinct points of view, and then to say that one of them – the objective one – has different width scopes. However, arguments that claim "we're more ontologically tidy" are often not persuasive, so we really base our position on what we see as its greater intuitive plausibility.

5. There is a lot of work to do on how this objectification/reification takes place. One particularly interesting idea is the notion of *psychological essentialism*. This is a hypothesized, but still poorly understood, psychological mechanism whereby ephemeral sense impressions are given rigidity and substance. It is thought that psychological essentialism might explain why we think that concepts have necessary and sufficient conditions for their application (or that objects have necessary and sufficient conditions defining them) even though they don't. The seminal work on this was done by Medin and Ortony. See their (1989). See also Murphy (2002).

6. We have not come close to giving the last word on the nature of subjective and objective points of view. This fascinating topic is deep and rich, and deserves its own monograph. Of course, we strongly recommend Nagel's works on the topic.

7. Nagel doesn't explicitly relate concepts to points of view in either his (1979) or his (1986), nor does he explicate the relation between beliefs and points of view. But Chalmers does. See his (2003).

8. This attending is a *shift in point of view* – a matter we will take up in detail in Chapter 6.

9. There are probably many different kinds of phenomenal concepts one can entertain if one is being appeared to red-squarely (to use a locution made famous by Wilfred Sellars). See Chalmers (2003), for a discussion of some of the more important of these.

10. There are deep issues about the relation between concepts and points of view that we are going to have to forgo. The question is: can different points of view attach to the same concept, or are points of view tied so intimately to concepts that concepts are partly individuated by their points of view and hence changing points of view is changing concepts? The first is the position that points of view can be bound to concepts like values can be bound to variables. In $X + Y = 7$, X and Y can take on different values but the variables X and Y remain the same. Do concepts and points of view behave like that? Or is a point of view part of what makes a given concept the concept that it is; do concepts have their points of view essentially? At this stage of consciousness research and given what cognitive science currently knows about concepts there can be no definitive answer. There is no settled, definitive theory of concepts that explains what makes a concept the concept that is (see Murphy 2002), let alone, how concepts and points of view are related – indeed, it is hard to find any psychological work on both points of view and concepts.

11. See Chapter 8 of his 1996 (a very interesting chapter). There, he considers a dual-aspect theory of consciousness based on information, where information has both a physical and a phenomen aspect.

12. The supervenience inference is blocked in the scientific dualism case for reasons that we discussed in this chapter, assuming you could get your hands on the relevant objective concepts. But that seems unlikely. We can't get any empirical data confirming that matter or information has a special, proto-phenomenal property. The argument for such a state of affairs has to be completely *a priori* and metaphysical. Such arguments are too speculative to give us a robust, objective concept. Hence, such arguments are too weak to support a scientific, yet dualistic, explanation of consciousness.

Chapter 4

1. Jackson (1982) sees quite a difference between his argument and Nagel's. In particular, Jackson thinks his own argument but not Nagel's causes problems for physicalism. We will not be concerned with either Nagelian or Jacksonian exegesis. We are interested in using their arguments to undermine faith in an eventual, useful science of consciousness.

2. We are using Nagel's and Jackson's arguments to make the case that a science of consciousness is not in the offing. We are not saying they are mysterians. For example, Nagel, at the end of his (1974), suggests that some sort of "objective phenomenology" might be created or developed. If this were possible, then perhaps a science could be developed using that. See, also, Nagel's (1998, 2002). But in truth, it is hard to tell what Nagel's position on the possibility of a science of consciousness is. However, Jackson may really be a mysterian. In his (1982), he argues for epiphenomenalism. This is, on some construals, arguably an anti-science-of-consciousness position because of its implication that consciousness is a causal

dead-end – as far as anyone knows, there are no causal dead ends postulated anywhere else in the natural world.

3. We have already shown in Part I that an explanatory science of consciousness is not in the cards. So when mysterians say there won't be one, they're right. But mysterians are claiming that there will *no* science of consciousness of any stripe. Since all their arguments focus on the existence of an unbridgeable explanatory gap, it is therefore open to us to claim that these mysterian arguments commit the fallacy of *non sequitur* by concluding that no science is possible from the fact that no explanatory science is possible. But claims that important arguments commit a fallacy aren't very illuminating. An in-depth analysis of Nagel's and Jackson's position and related mysterian arguments is what is needed to lay to rest any plausibility for the mysterian view.

4. We relish the irony. We are trying to make it intuitively plausible that a science of consciousness will come along that is not intuitively plausible.

5. Compare Chalmers (1996: Chapter 6). There he develops the rudiments of a nonreductive theory of consciousness by focusing on the coherence between consciousness and cognition, which is similar to our experience and description.

Chapter 5

1. For a good, extended analysis of the neural correlates of consciousness and the state of play in locating them (as well as an example of the kind of science we are advocating), see Koch (2004).

2. This section draws on Flohr (1992, 1995a, 1995b) and Flohr et al. (1998).

3. We take this term from Wilson (1999).

4. Indeed, we take this last point to give us one good reason to be anti-realists (of a sort) about consciousness studies. Nothing about our theories is going to tell us whether materialism or dualism is true; hence, we should stop seeking to answer that question and instead focus on what we can do. What we can do is develop a science of consciousness that ignores those metaphysical issues.

Chapter 6

1. Of course, the fight continues. See for example, Daniel Wegner's account of conscious will as an illusion, in his book *The Illusion of Conscious Will* (2002). A feeling of ennui sets in. Sisyphus smiles. He has company.

2. The belief need not be conscious. But it has to be possible to bring it to consciousness.

3. Nagel, of course considers the self in his works on the subjective and objective. But he focuses on the philosophical problem of the self as one's perspective shifts from subjective to objective. See his (1979, 1986). One might try to argue that since Nagel focuses on specific

kinds of viewpoint change, namely, between subjective and objective viewpoints, he gets a subject having those points of view for free. But this would be to misunderstand our third necessary condition. We are saying that changing points of view *simpliciter* requires a subject. This is both more general than Nagel's version (*all* viewpoint change requires a subject), and more specific, since it is the maintenance of continuity through change that requires the subject.

4. Compare Chalmers (1996: 196ff.).

5. If we assume that one is conscious of the viewpoint change itself (say, after having it pointed out by Nagel) then we require that one be conscious of a maintained referent; that is, one referent, from different points of view, exists in consciousness. One needn't be conscious *that* the referent is being maintained. That is, one needn't be conscious that the referent doesn't change. It is sufficient that there is one maintained referent in consciousness. Of course, this doesn't entail that out in the world, there is only one referent. A maintained referent is a mental object, the proximate, mental side of a reference-fixing mechanism; a referent in the world need not be. Being conscious of a maintained referent might be (part of) what explains the fact that we are conscious of ourselves as selves. Necessary condition 3), above, guarantees that there is a subject who is the locus of changes of points of view. But once consciousness is added in, that might be sufficient for guaranteeing that the self in question is a conscious self. For, if you are conscious of your changing points of view then it might be that you are thereby conscious of being a self whose points of view are changing. Fortunately, we don't need to decide this difficult issue here.

6. We aren't saying no branch of philosophy makes progress. Certain aspects of political philosophy and feminist philosophy of science seem to have enhanced our understanding of the world, as have branches of the philosophy of mind concerned with cognition and representation.

7. We don't want to get sidetracked into aesthetics, but we do want to point out that it does not follow from our view that humans cannot judge one work of art to be better than another. Nor does it follow from our claims that one work of art cannot *be* better than another. Works of art can and do vary in their ability to accomplish their goals, which presumably is to affect humans emotionally and intellectually in one way or another. Art doesn't make progress, but within any given milieu and genre, the success of various works can be compared.

8. See Nagel (1979, 1986).

Appendix

1. Much of the argument in this and the next three sections was developed by Anthony Gillies, see Dietrich and Gillies (2001).

2. Actually, as Chalmers notes, this can't be the whole story, for it is possible that there is a universe physically identical to ours, but which also contains additional nonphysical stuff not present in our own world – angels, ectoplasm, ghosts, and the like. If these angels follow biological laws, say, and differentially reproduce and evolve, then biology might not logically

supervene on the physical. But, as Chalmers says, "we certainly want to say that biological properties are [logically] supervenient on physical properties, at least in this world ... Intuitively, it seems undesirable for the mere logical possibility of the angel world to stand in the way of the determination of biological properties by physical properties in our own world" (1996:39). He concludes that we need to restrict our notion of logical supervenience a wee bit. He offers two restrictions. First, we should make supervenience into a thesis about our universe (or more generally, about particular universes). So A-facts logically supervene on B-facts if in any possible world with B-facts, at least the A-facts will be true. Any additional, extra A-facts (the existence of angels, say) will not count against the supervenience relation. Second, because this restriction doesn't help with the supervenience of certain general facts in our world – that there is no such thing as ectoplasm, for example – Chalmers disallows negative claims (this would also include universal statements such as all kangaroos are mammals). "Supervenience theses should apply only to positive facts and properties, those that cannot be negated simply by enlarging a world" (1996:40).

3. Some have balked at this move. They claim that it is an instance of the problem of other minds and that this problem is solved, or at least not a serious problem. Again, however, if the problem were solved or widely regarded as not serious, there would be near unanimity about this, as in mathematics or science. But there is not. It is not crazy, nor even radical, to claim that the problem of other minds has not been solved, nor is it crazy to regard it as a serious difficulty. Of course, in our daily lives, we like everyone else, find it impossible to sustain worries that others have phenomenal states, but ordinary lives are not the stuff of philosophy; seeking deeper "truths" is. And the deep truth here is that, for all we really know, where "really know" means being philosophically certain, all others but ourselves are zombies. Indeed, this understates the case. Properly stated, the claim is that for all I know (where "I" can function like a variable for whoever reads this sentence), everyone but me is a zombie. Is our definition of "really know" too strong? That is just the problem of other minds all over again. Within some parts of philosophy, the answer seems to be "No." But in day-to-day life, the answer is clearly "Yes." Indeed, even wondering whether this day-to-day answer is relevant to the former, philosophical answer, is an aspect of the problem of other minds.

4. One might think that, upon analysis, it could turn out that consciousness is such and such a brain process. Neither we nor Chalmers are begging any questions here. If it should turn our that consciousness is some brain process, then this would not be the secondary intension of consciousness, for we can only pick out consciousness via first-person access. And given this kind of access, the primary and secondary intensions coincide.

5. Chalmers's definition for a conceivable sentence is one that is true at *all* conceivable worlds. If he means this as it is written, it has the consequence that all contingently true sentences are inconceivable! For the sake of charitability, we will regard '1-conceivability' as conceivability under the primary intensions of the terms involved, leaving 'conceivable' in its standard, intuitive sense.

6. Levine's (1993) paper also questions armchair metaphysics. Nagel, too, arguably also questions armchair metaphysics. He says "Perhaps there could not actually be such robots [which behaved like people though they experienced nothing]. Perhaps anything complex enough to behave like a person would have experience. But that, if true, is a fact which cannot be discovered merely by analyzing the concept of experience" (1974:2f.). It is somewhat plausible that by "actually," Nagel is not referring only to the actual world. Still, it does seem as if he has run actuality and possibility together. Nevertheless, the sentiment is clear: mere conceptual analysis is not up to the job of settling these complex questions involving whether or not the phenomenal and the physical realm can pull apart.

7. In Chapter 3 of this 1996 book, Chalmers says "So let us consider my zombie twin. This creature is molecule for molecule identical to me, and identical in all the *low-level* properties postulated by a completed physics, but he lacks conscious experience entirely" (p. 94) (our emphasis). It is clear that, in conceiving of his zombie twin, Chalmers has already sundered the physical from the phenomenal. Deriving dualism, then, is straightforward.

8. Of course, Chalmers doesn't think this inference is invalid. See Chalmers and Jackson (2001). But see Balog (1999).

9. Further problems with these modal, zombie intuitions can be seen by considering the problem that for Chalmers phenomenal judgments about consciousness don't require consciousness. Consciousness is *irrelevant* to judgments about consciousness. This is because judgments about consciousness are strictly psychological phenomena. Hence, our "zombie twins" will make such judgments; in particular, they will judge that they are conscious. This is deeply troubling, for, according to dualists, the zombies aren't conscious.

References

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Baars, B. & Newman, J. (1994). A neurobiological interpretation of global workspace theory. In A. Revonsuo & M. Kampinen (Eds.), *Consciousness in Philosophy and Cognitive Neuroscience* (pp. 211–226). Hillsdale, NJ: Lawrence Erlbaum.

Balog, K. (1999). Conceivability, possibility, and the mind-body problem. *Philosophical Review, 108*, 497–528.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22* (4), 577–660.

Block, N. (1997). On a confusion about a function of consciousness. In N. Block, O. Flanagan, & G. Guzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.

Block, N. & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review, 108*, 1–46.

Brandon, R. N. (1982). The levels of selection. In *PSA 1982*, Vol. I (pp. 315–323). East Lansing, MI: Philosophy of Science Association.

Cartwright, N. (1970). Causal laws and effective strategies. *Nous, 13*, 419–437.

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2* (3), 200–219.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chalmers, D. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New Philosophical Essays*. Oxford: OUP.

Chalmers, D. & Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review, 110*, 315–361.

Churchland, P. M. (1984). *Matter and Consciousness*. Cambridge, MA: MIT Press.

Churchland, P. M. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA: MIT Press.

Crick, F. & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in Neuroscience, 2*, 263–275.

Davidson, D. (1970/1993). Mental events. In Foster & Swanson (Eds.), *Experience and Theory*. London: Duckworth, 1970; and Kim, *Supervenience and Mind*, Cambridge: Cambridge University Press, 1993.

Deacon, T. (1997). *The Symbolic Species*. New York: W. W. Norton.

Dehaene, S. (Ed.). (2001). *The Cognitive Neuroscience of Consciousness*. Amsterdam and Cambridge: Elsevier and MIT Press.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown, and Company.

Dietrich, E. (2001). Concepts: Fodor's little semantic BBs of thought – A critical look at Fodor's theory of concepts. *J. Experi. and Theor. AI, 3* (2), 89–94.

Dietrich, E. & Gillies, A. (2001). Consciousness and the limits of our imaginations. *Synthese, 126* (3), 361–381.

Dietrich, E. & Hardcastle, V. (2002). A Connecticut Yalie in King Descartes' Court: A review of *Mind and Mechanism* by Drew McDermott. *The Newsletter of the Cognitive Science Society, 22* (2), June. http://www.cognitivesciencesociety.org/newsletter/June02/index.html

Dietrich, E. & Markman, A. (2000). Cognitive dynamics: Computation and representation regained. In E. Dietrich & A. Markman (Eds.), *Cognitive Dynamics: Conceptual Change in Humans and Machines*. Mahwah, NJ: Lawrence Erlbaum.

Eells, E. (1988). Probabilistic causal laws. In B. Skyrms & W. I. Harper (Eds.), *Causation, Chance and Credence*, Vol. 1 (pp. 109–133). New York: Reidel.

Feigl, E. (1967). *The 'mental' and the 'physical'*. Minneapolis: University of Minnesota Press.

Flanagan, O. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Flohr, H. (1992). Qualia and brain processes. In A. Beckerman, H. Flohr, & J. Kim (Eds.), *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism* (pp. 220–238). New York: Walter de Gruyter.

Flohr, H. (1995a). An information processing theory of anesthesia. *Neuropsychologia, 33*, 1169–1180.

Flohr, H. (1995b). Sensations and brain processes. *Behavioral Brain Research, 71*, 157–161.

Flohr, H., Glade, U., & Motzko, D. (1998). The role of the NMDA synapse in general anesthesia. *Toxicology Letters, 100–101*, 23–29.

Fodor, J. (1983). *The Modularity of Mind: An Essay in Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor, J. (1994). *The Elm and the Expert*. Cambridge, MA: MIT Press.

Fodor, J. (1998a). *In Critical Condition*. Cambridge, MA: MIT Press.

Fodor, J. (1998b). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.

Gazzaniga, M., Fendrich, R. F., & Wessinger, C. M. (1994). Blindsight reconsidered. *Current Directions in Psychological Science, 3*, 93–96.

Giesy, Courtney & Dietrich, E. (2001). Review of *Concepts: Where cognitive science went wrong*, by Jerry Fodor. *Newsletter of the Cognitive Science Society*. http://www.cognitivesciencesociety.org/newsletter/June01/FodRev.html

Hammeroff, S. (1994). Quantum coherence in microtubules: A neural basis for emergent consciousness. *Journal of Consciousness Studies, 1*, 91–118.

Hardcastle, V. G. (1991). Partitions, probabilistic laws, and Simpson's paradox. *Synthese, 86*, 209–228.

Hardcastle, V. G. (1995). *Locating Consciousness*. Philadelphia: John Benjamins Press.

Hardcastle, V. G. (1998). On the matter of mental causation. *Philosophy and Phenomenological Review*.

Hardin, C. L. (1988). *Color for Philosophers: Unweaving the Rainbow*. New York: Hackett.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Hintikka, J. (1993). On proper (popper?) and improper uses of information in epistemology. *Theoria, LIX*, 158–165.

Hubbard, T. L. (1996). The importance of a consideration of qualia to imagery and cognition. *Consciousness and Cognition, 5*, 327–358.

Jack, A. I. & Shallice, T. (2001). Introspective physicalism as an approach to the science of consciousensss. In S. Dehaene (Ed.).

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly, 32*, 127–136.

Keil, F. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. Premack (Eds.), (pp. 234–267).

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.

Kim, J. (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.

Koch, C. (2004). *The Quest for Consciousness*. Roberts: Englewood, Colorado.

Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and Brain*. Cambridge, MA: MIT Press.

Kripke, S. (1972). Naming and necessity. In G. Harman & D. Davidson (Eds.), *The Semantics of Natural Language*. Dordrecht: Reidel.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lerner, R. & Damon, W. (Eds.). (2000). *Handbook of Child Psychology: Theoretical Models of Human Development*. New York: Wiley.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*, 354–361.

Levine, J. (1993). On leaving out what it is like. In M. Davies & G. Humphreys (Eds.), *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.

Lewis, D. (1986). *On the Plurality of Worlds*, Oxford: Blackwell.

Margolis, E. & Laurence, S. (Eds.). (1999). *Concepts: Core readings*. Cambridge, MA: MIT Press.

Markman, A. & Dietrich, E. (2000). In defense of representations. *Cognitive Psychology, 40*, 138–171.

McDermott, D. (2001). *Mind and Mechanism*. Cambridge, MA: MIT Press.

McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.

McGinn, C. (1989). "Can we solve the mind-body problem?" *Mind, 93*, 349–366.

McGinn, C. (1991). *The Problem of Consciousness*. Oxford: Blackwell.

McGinn, C. (1993). *Problems in Philosophy: The Limits of Inquiry*. Oxford: Blackwell.

McGinn, C. (1999). *The Mysterious Flame*. New York: Basic Books.

Medin, D. & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 179–195). Cambridge: Cambridge Univ. Press.

Mellor, D. H. (1977). Conscious belief. *Proceedings of the Aristotelian Society, 78*, 87–101.

Murphy, G. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*, 435–450. Reprinted in Nagel (1979).

Nagel, T. (1979). Subjective and objective. In *Mortal Questions* (pp. 196–213). Cambridge, UK: Cambridge University Press.

Nagel, T. (1986). *The View From Nowhere*. New York: Oxford University Press.

Nagel, T. (1998). Conceiving the impossible and the mind-body problem. *Philosophy, 73* (285), 337–352.
Nagel, T. (2002). *Concealment and Exposure and Other Essays*. New York: Oxford University Press.
O'Leary-Hawthorne, J. (1996). The epistemology of possible worlds. *Philosophical Studies, 84*, 183–202.
Perkins, M. (1971). Sentience. *Journal of Philosophy, 68*, 329–337.
Port, R. & van Gelder, T. (Eds.). (1995). *Mind as Motion*. Cambridge, MA: MIT Press.
Priest, G. (2002). *Beyond the Limits of Thought*. Oxford, UK: Oxford University Press.
Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (Ed.), *Language, Mind & Knowledge* (pp. 131–193). Minneapolis: University of Minnesota Press.
Raffman, D. (1993). *Language, Music and Mind*. Cambridge, MA: MIT Press.
Russell, B. (1912/1959). *The Problems of Philosophy*. Oxford: Oxford University Press.
Russell, B. (1961). *Religion and Science*. New York: Oxford University Press.
Salmon, N., (1996). Trans-world identification and stipulation. *Philosophical Studies, 84*, 203–223.
Salmon, W. (1971). *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.
Skarda, C. A. & Freeman, W. J. (1990). Chaos and the new science of the brain. *Concepts in Neuroscience, 1*, 275–285.
Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*, 417–424.
Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Smolin, L. (1995). *The Life if the Cosmos: A New View of Cosmology, Particle Physics, and the Meaning of Quantum Physics*. New York: Crown.
Sperber, D., Premack, D., & Premack, A. (Eds.). (1995). *Causal Cognition: A multidisciplinary Debate. Symposia of the Fyssen Foundation*. New York: Carendon Press.
Taylor, J. (1998). Cortical activity and the explanatory gap. *Consciousness and Cognition, 7* (2), 109–148.
Taylor, J. (2001). *The Race for Consciousness*. Cambridge, MA: MIT Press.
Thelen, E., & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.
von der Malsburg, C. (1981). The correlation theory of brain function. Internal Report 81-2. Max-Planck Institute for Biophysical Chemistry.
Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
Wilson, R. A. (1999). The individual in biology and psychology. In V. G. Hardcastle (Ed.), *Biology Meets Psychology: Conjectures, Connections, Constraints*. Cambridge, MA: MIT Press.
Wimsatt, W. C. (1984). Reductive explanation: A functional account. In E. Sober (Ed.), *Conceptual Issues and Their Biases in the Units of Selection Controversy* (pp. 142–183). Cambridge, MA: MIT.
Weiskrantz, L. & Cowey, A. (1967). A comparison of the effects of striate cortex and retinal lesions on visual acuity in the monkey. *Science, 155*, 104–106.
Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences, 7* (5), 214–218.

Index

In the series *Advances in Consciousness Research* the following titles have been published thus far or are scheduled for publication:

3 **JIBU, Mari and Kunio YASUE:** Quantum Brain Dynamics and Consciousness. An introduction. 1995. xvi, 244 pp.

4 **HARDCASTLE, Valerie Gray:** Locating Consciousness. 1995. xviii, 266 pp.

5 **STUBENBERG, Leopold:** Consciousness and Qualia. 1998. x, 368 pp.

6 **GENNARO, Rocco J.:** Consciousness and Self-Consciousness. A defense of the higher-order thought theory of consciousness. 1996. x, 220 pp.

7 **MAC CORMAC, Earl and Maxim I. STAMENOV (eds.):** Fractals of Brain, Fractals of Mind. In search of a symmetry bond. 1996. x, 359 pp.

8 **GROSSENBACHER, Peter G. (ed.):** Finding Consciousness in the Brain. A neurocognitive approach. 2001. xvi, 326 pp.

9 **Ó NUALLÁIN, Seán, Paul Mc KEVITT and Eoghan Mac AOGÁIN (eds.):** Two Sciences of Mind. Readings in cognitive science and consciousness. 1997. xii, 490 pp.

10 **NEWTON, Natika:** Foundations of Understanding. 1996. x, 211 pp.

11 **PYLKKÖ, Pauli:** The Aconceptual Mind. Heideggerian themes in holistic naturalism. 1998. xxvi, 297 pp.

12 **STAMENOV, Maxim I. (ed.):** Language Structure, Discourse and the Access to Consciousness. 1997. xii, 364 pp.

13 **VELMANS, Max (ed.):** Investigating Phenomenal Consciousness. New methodologies and maps. 2000. xii, 381 pp.

14 **SHEETS-JOHNSTONE, Maxine:** The Primacy of Movement. 1999. xxxiv, 584 pp.

15 **CHALLIS, Bradford H. and Boris M. VELICHKOVSKY (eds.):** Stratification in Cognition and Consciousness. 1999. viii, 293 pp.

16 **ELLIS, Ralph D. and Natika NEWTON (eds.):** The Caldron of Consciousness. Motivation, affect and self-organization — An anthology. 2000. xxii, 276 pp.

17 **HUTTO, Daniel D.:** The Presence of Mind. 1999. xiv, 252 pp.

18 **PALMER, Gary B. and Debra J. OCCHI (eds.):** Languages of Sentiment. Cultural constructions of emotional substrates. 1999. vi, 272 pp.

19 **DAUTENHAHN, Kerstin (ed.):** Human Cognition and Social Agent Technology. 2000. xxiv, 448 pp.

20 **KUNZENDORF, Robert G. and Benjamin WALLACE (eds.):** Individual Differences in Conscious Experience. 2000. xii, 412 pp.

21 **HUTTO, Daniel D.:** Beyond Physicalism. 2000. xvi, 306 pp.

22 **ROSSETTI, Yves and Antti REVONSUO (eds.):** Beyond Dissociation. Interaction between dissociated implicit and explicit processing. 2000. x, 372 pp.

23 **ZAHAVI, Dan (ed.):** Exploring the Self. Philosophical and psychopathological perspectives on self-experience. 2000. viii, 301 pp.

24 **ROVEE-COLLIER, Carolyn, Harlene HAYNE and Michael COLOMBO:** The Development of Implicit and Explicit Memory. 2000. x, 324 pp.

25 **BACHMANN, Talis:** Microgenetic Approach to the Conscious Mind. 2000. xiv, 300 pp.

26 **Ó NUALLÁIN, Seán (ed.):** Spatial Cognition. Foundations and applications. 2000. xvi, 366 pp.

27 **GILLETT, Grant R. and John McMILLAN:** Consciousness and Intentionality. 2001. x, 265 pp.

28 **ZACHAR, Peter:** Psychological Concepts and Biological Psychiatry. A philosophical analysis. 2000. xx, 342 pp.

29 **VAN LOOCKE, Philip (ed.):** The Physical Nature of Consciousness. 2001. viii, 321 pp.

30 **BROOK, Andrew and Richard C. DEVIDI (eds.):** Self-Reference and Self-Awareness. 2001. viii, 277 pp.

31 **RAKOVER, Sam S. and Baruch CAHLON:** Face Recognition. Cognitive and computational processes 2001. x, 306 pp.

32 **VITIELLO, Giuseppe:** My Double Unveiled. The dissipative quantum model of brain. 2001. xvi, 163 pp.

33 **YASUE, Kunio, Mari JIBU and Tarcisio DELLA SENTA (eds.):** No Matter, Never Mind. Proceedings of Toward a Science of Consciousness: Fundamental approaches, Tokyo 1999. 2002. xvi, 391 pp.

34 **FETZER, James H. (ed.):** Consciousness Evolving. 2002. xx, 253 pp.

35 **Mc KEVITT, Paul, Seán Ó NUALLÁIN and Conn MULVIHILL (eds.):** Language, Vision and Music. Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, 1999. 2002. xii, 433 pp.

36 **PERRY, Elaine, Heather ASHTON and Allan H. YOUNG (eds.):** Neurochemistry of Consciousness. Neurotransmitters in mind. With a foreword by Susan Greenfield. 2002. xii, 344 pp.

37 **PYLKKÄNEN, Paavo and Tere VADÉN (eds.):** Dimensions of Conscious Experience. 2001. xiv, 209 pp.

38 **SALZARULO, Piero and Gianluca FICCA (eds.):** Awakening and Sleep–Wake Cycle Across Development. 2002. vi, 283 pp.

39 **BARTSCH, Renate:** Consciousness Emerging. The dynamics of perception, imagination, action, memory, thought, and language. 2002. x, 258 pp.

40 **MANDLER, George:** Consciousness Recovered. Psychological functions and origins of conscious thought. 2002. xii, 142 pp.

41 **ALBERTAZZI, Liliana (ed.):** Unfolding Perceptual Continua. 2002. vi, 296 pp.

42 **STAMENOV, Maxim I. and Vittorio GALLESE (eds.):** Mirror Neurons and the Evolution of Brain and Language. 2002. viii, 392 pp.

43 **DEPRAZ, Nathalie, Francisco J. VARELA and Pierre VERMERSCH:** On Becoming Aware. A pragmatics of experiencing. 2003. viii, 283 pp.

44 **MOORE, Simon C. and Mike OAKSFORD (eds.):** Emotional Cognition. From brain to behaviour. 2002. vi, 350 pp.

45 **DOKIC, Jérôme and Joëlle PROUST (eds.):** Simulation and Knowledge of Action. 2002. xxii, 271 pp.

46 **MATEAS, Michael and Phoebe SENGERS (eds.):** Narrative Intelligence. 2003. viii, 342 pp.

47 **COOK, Norman D.:** Tone of Voice and Mind. The connections between intonation, emotion, cognition and consciousness. 2002. x, 293 pp.

48 **JIMÉNEZ, Luis (ed.):** Attention and Implicit Learning. 2003. x, 385 pp.

49 **OSAKA, Naoyuki (ed.):** Neural Basis of Consciousness. 2003. viii, 227 pp.

50 **GLOBUS, Gordon G.:** Quantum Closures and Disclosures. Thinking-together postphenomenology and quantum brain dynamics. 2003. xxii, 200 pp.

51 **DROEGE, Paula:** Caging the Beast. A theory of sensory consciousness. 2003. x, 183 pp.

52 **NORTHOFF, Georg:** Philosophy of the Brain. The brain problem. 2004. x, 433 pp.

53 **HATWELL, Yvette, Arlette STRERI and Edouard GENTAZ (eds.):** Touching for Knowing. Cognitive psychology of haptic manual perception. 2003. x, 322 pp.

54 **BEAUREGARD, Mario (ed.):** Consciousness, Emotional Self-Regulation and the Brain. 2004. xii, 294 pp.

55 **PERUZZI, Alberto (ed.):** Mind and Causality. 2004. xiv, 235 pp.

56 **GENNARO, Rocco J. (ed.):** Higher-Order Theories of Consciousness. An Anthology. 2004. xii, 371 pp.

57 **WILDGEN, Wolfgang:** The Evolution of Human Language. Scenarios, principles, and cultural dynamics. 2004. xii, 240 pp.

58 **GLOBUS, Gordon G., Karl H. PRIBRAM and Giuseppe VITIELLO (eds.):** Brain and Being. At the boundary between science, philosophy, language and arts. 2004. xii, 350 pp.

59 **ZAHAVI, Dan, Thor GRÜNBAUM and Josef PARNAS (eds.):** The Structure and Development of Self-Consciousness. Interdisciplinary perspectives. 2004. xiv, 162 pp.

60 **DIETRICH, Eric and Valerie Gray HARDCASTLE:** Sisyphus's Boulder. Consciousness and the limits of the knowable. 2005. xii, 133 pp.

61 **ELLIS, Ralph D.:** Curious Emotions. Roots of consciousness and personality in motivated action. vii, 233 pp. + index. *Expected Spring 2005*

62 **DE PREESTER, Helena and Veroniek KNOCKAERT (eds.):** Body Image and Body Schema. Interdisciplinary perspectives on the body. ca. 360 pp. *Expected Summer 2005*