

# THE CHARACTER OF CONSCIOUSNESS

---

---

DAVID J. CHALMERS

**OXFORD**  
UNIVERSITY PRESS

2010

# CONTENTS

---

---

Introduction	xii
<b>I. The Problems of Consciousness</b>	
1. Facing Up to the Problem of Consciousness	3
<i>Afterword: From "Moving Forward on the Problem of Consciousness"</i>	28
<b>II. The Science of Consciousness</b>	
2. How Can We Construct a Science of Consciousness?	37
<i>Afterword: First-Person Data and First-Person Science</i>	52
3. What Is a Neural Correlate of Consciousness?	59
4. On the Search for the Neural Correlate of Consciousness	91
<b>III. The Metaphysics of Consciousness</b>	
5. Consciousness and Its Place in Nature	103
6. The Two-Dimensional Argument against Materialism	141
<i>Afterword: Other Anti-Materialist Arguments</i>	192
7. Conceptual Analysis and Reductive Explanation (with Frank Jackson)	207
<b>IV. Concepts of Consciousness</b>	
8. The Content of Phenomenal Concepts	251
9. The Epistemology of Phenomenal Belief	277
10. Phenomenal Concepts and the Explanatory Gap	305

## V. The Contents of Consciousness

- |     |  |     |
|-----|--|-----|
| 11. | The Representational Character of Experience                 | 339 |
|     | <i>Afterword: The Two-Dimensional Contents of Perception</i> | 372 |
| 12. | Perception and the Fall from Eden                            | 381 |
| 13. | The Matrix as Metaphysics                                    | 455 |
|     | <i>Afterword: Philosophical Notes</i>                        | 479 |

## VI. The Unity of Consciousness

- |     |  |     |
|-----|--|-----|
| 14. | What Is the Unity of Consciousness? (with Tim Bayne) | 497 |
|     | Appendix: Two-Dimensional Semantics                  | 541 |
|     | Bibliography   | 569 |
|     | Subject Index  | 587 |
|     | Name Index   | 593 |

## INTRODUCTION

---

---

What is consciousness? How can it be explained? Can there be a science of consciousness? What is the neural basis of consciousness? What is the place of consciousness in nature? Is consciousness physical or nonphysical? How do we know about consciousness? How do we think about consciousness? What are the contents of consciousness? How does consciousness relate to the external world? What is the unity of consciousness?

We can think of these questions as limning a few dimensions of the character of consciousness. Consciousness is an extraordinary and multifaceted phenomenon whose character can be approached from many different directions. It has a phenomenological and a neurobiological character. It has a metaphysical and an epistemological character. It has a perceptual and a cognitive character. It has a unified and a differentiated character. And it has many further sorts of character.

We will not understand consciousness by studying its character on just one of these dimensions. Studying the phenomenology or the neurobiology of consciousness alone may tell us a great deal, as might studying the metaphysics or the epistemology. The perceptual and cognitive aspects of consciousness pose huge challenges in their own right. But ultimately we must approach consciousness from all of these directions.

In this book, I address all of these issues about the character of consciousness and a number of others. It is not the last word (or even my last word) on any of them. There are many aspects of the character of consciousness that it does not address at all. Still, I hope that it provides a unified picture of many aspects of consciousness that repays attention.

The chapters of this book were first written as separate articles, so one might think that the book is bound to be fragmented. I have tried to structure and rework it in such a way that it works as a whole, however. Later chapters build on ideas put forward in earlier chapters, and there are many common themes throughout. In principle one could read the book from start to finish as if following a narrative thread. The book is long enough

that perhaps this is too much to expect, but there are also many subbooks within it, some of which I will explain. I have added afterwords to some of the chapters, as well as new footnotes (marked with an asterisk). I have shifted some of the material and cut out the more blatant acts of repetition, although occasional repetition has survived, in part to assist readers who may not be reading straight through.

The book is structured as follows. Chapter 1 introduces the problems of consciousness in an accessible way. Chapters 2–4 address the science of consciousness by developing a positive picture of how the science works in light of the problems. Chapters 5–7 address the metaphysics of consciousness, arguing in detail against materialist views and for a view on which consciousness is irreducible. Chapters 8–10 address thought about consciousness and the epistemology of consciousness, developing an account of the concepts we use to think about consciousness and the distinctive knowledge that we have of consciousness. Chapters 11–13 address perceptual consciousness and the way it represents the external world. Chapter 14 addresses the unity of consciousness. The appendix gives an outline of the two-dimensional framework that plays a central role in a number of the more technical chapters.

Although the book contains some technical material, it also contains much that is intended to be highly accessible. Chapters 1–4 and 13 were written with a general audience in mind, and chapters 5 and 14 should be accessible to a reasonably broad audience. Chapters 6–12 are more technical and are likely to appeal mainly to philosophers or to those who are willing to work hard. To someone without much background in philosophy, I suggest skipping these chapters on a first reading. If pressed for time, start with chapters 1, 2, and 13, which have perhaps the broadest appeal. Likewise, for those especially interested in the science of consciousness, chapters 1–4 are the places to start, with perhaps some material of interest in chapters 13 and 14.

There are also many paths through the book for philosophers. Those especially interested in the mind-body problem might focus on chapters 1, 5–7, and 10 (with some relevant material in 2, 8, and 9). Those interested in issues about language, content, and concepts might focus on chapters 6–8 and 11–13, as well as the appendix. Those interested in epistemology might focus first on chapters 9 and 13, with relevant material also in 2, 4, 6, and 7. Those interested in phenomenology might find material of interest in chapters 11, 12, and 14. I urge philosophers not to skip chapter 13: the topic might seem frivolous, but the philosophical issues go as deep as in any chapter in the book.

This book could be considered a sort of sequel to my earlier book on consciousness, *The Conscious Mind* (*TCM*). That book received far more

attention than I could reasonably have expected, for which I am enormously grateful. At the same time, it is very far from a perfect work (most of it was written as my PhD dissertation after four years of studying philosophy), and there is much about it that I would change if I were writing it now. There are also many relevant issues that it simply does not address. I would like to think that in subsequent years, I have come to understand a number of important issues better than I did then. In this book, I aim to flesh out a picture of consciousness that is clearer, fuller, and more adequate than the picture in *The Conscious Mind*.

The picture in this book is largely consistent with the picture in *The Conscious Mind*. I have not had an enormous change of mind since then, though there are some medium-sized changes: for example, I am somewhat less sympathetic to epiphenomenalism than I was then and somewhat more sympathetic to drawing close connections between consciousness and intentionality. A few chapters present arguments along the same lines as those in that book. In particular, chapters 1 and 5–7 cover the same sort of ground as the early chapters of *TCM*, in what I hope is a better fashion. In some of those chapters I have also responded to various critics of *TCM*. The chapters in sections II and IV address issues regarding the science of consciousness and concepts of consciousness that are touched on far too briefly in *TCM*, while sections V and VI move in new directions that are not explored at all there (those who want to explore entirely different issues could start here). I have not presupposed any knowledge of the earlier book; this book can stand alone.

One regret concerning *The Conscious Mind* is that the book has become known especially for a negative thesis, the thesis that consciousness is not physical. Perhaps this was to be expected, but my central goal in studying consciousness has always been positive rather than negative: I would like to find a theory of consciousness that works. Negative arguments are just a step on the way. So while this book contains its share of negative arguments (mainly in section III and to some extent in chapters 1 and 10), I have tried to put more focus on the positive. Chapters 2–4 outline a positive picture of how the science of consciousness can proceed (and is proceeding) even once the distinctive problems of consciousness are fully recognized. The second halves of chapters 1 and 5 sketch speculative positive pictures of the metaphysics of consciousness; chapters 8 and 9 try to tell a positive story about phenomenal concepts and the epistemology of consciousness; and the chapters in sections V and VI develop a positive picture of perceptual consciousness, its relation to the external world, and the unity of consciousness.

It is worth noting that many elements of this positive picture can be accepted by those who are not convinced by the negative arguments. Apart from chapters 1, 4–6, and 10, most of this book is quite compatible with various sorts of materialism. At the same time, there is at least a nonreductive spirit to many aspects of the picture. One might think of the book as a whole as an attempt to articulate a reasonably unified, broadly nonreductive vision of the character of consciousness.

In the remainder of this introduction I introduce each of the chapters in more detail, sketching some context and some informal background. I have permitted myself a few more personal remarks than elsewhere in the book. Readers can read this material or not as they like. The material can also be read before or after reading the chapters themselves.

Chapter 1 introduces the distinction between the “hard problem” and the “easy problems” of consciousness. The easy problems are those of explaining cognitive and behavioral functions such as discrimination, integration, and verbal report. The hard problem is that of explaining conscious experience. Where the easy problems are concerned, it suffices to explain how a function is performed, and to do this it suffices to specify an appropriate neural or computational mechanism. But where the hard problem is concerned, explaining cognitive and behavioral functions always leaves a further open question: why is the performance of these functions accompanied by experience? Because of this, the standard reductive methods of neuroscience and cognitive science that work for the easy problems do not work for the hard problem. I argue that this problem applies to any reductive explanation. In principle, these can explain only structure and function, and explaining conscious experience requires more than explaining structure and function. If this is right, there can be no wholly reductive explanation of consciousness. I go on to sketch an alternative nonreductive approach in which consciousness is taken as fundamental, and I sketch the outlines of a speculative theory of that sort.

I first presented this material in a talk at the first Tucson conference on the science of consciousness in April 1994. Although there is nothing especially original about the idea that consciousness poses a hard problem, I was in the right place at the right time. At the conference, the talk received an overwhelming and entirely unexpected reaction. Shortly afterward, I was invited to write up versions of this material for both *Scientific American* and the *Journal of Consciousness Studies*. The latter journal published four special issues containing a keynote article by me, around twenty-five responses, and a response by me in turn, all of which were later collected into the 1997 book titled *Explaining Consciousness: The Hard Problem*

(edited by Jonathan Shear). The chapter here is a version of the original article in that journal.

As an afterword to chapter 1, I have also included an excerpt from my lengthy reply to the responses, “Moving Forward on the Problem of Consciousness.” That article is too long to include here in full (it is available online), and much of it covers ground that is covered elsewhere in this book. But I have included a response to Daniel Dennett, both here and in the afterword to chapter 2, as his brand of highly deflationary materialism (the “type-A” materialism of chapter 5) is not discussed at length elsewhere in the book. I know that the debate between the deflationary and inflationary views of consciousness is of interest to readers both inside and outside philosophy, where the argument between the likes of Dennett and the likes of me is sometimes cast as an ideological battle. I learned long ago that I am not much of an ideological crusader: I do not believe that the world will be a hugely better place if everyone accepts my views, and I am only occasionally inclined to rhetorical extremes. Still, I think that one has a responsibility to the views that one has put forward, and the back-and-forth can at least shed some light on the underlying issues. So I have done my best to indicate how I see the core issues here.

I am fond of the original article on the hard problem in part because it makes the case against reductive theories of consciousness without any technicalities or far-out philosophical thought experiments. I have no objection to technicalities and thought experiments, which play a central role in the arguments of *The Conscious Mind*, but I think that the argument here is just as effective. There is a sense in which the argument here, which turns on simple issues about explanation, is more fundamental than conceivability arguments involving zombies, epistemological arguments involving Mary in her black-and-white room, and the like. And certainly it works better for an audience outside philosophy. It is sometimes supposed that nonreductive arguments turn essentially on these thought experiments, but this is just wrong. In fact, in chapter 5 and elsewhere, I suggest that the thought experiments turn essentially on points about structure and function that are close to the central points in this chapter. I see the thought experiments as a useful technical device for making the arguments more formal and more analyzable (as well as more vivid), and as a result they have produced much interesting and useful philosophical literature. Still, many of the common responses to those thought experiments have no clear application as a response to the simple arguments here (the phenomenal concept strategy discussed in chapter 10 is an exception, among others). Although the material in this chapter has been very widely cited, I have seen many fewer serious replies to its line of argument than to



the line in *The Conscious Mind*. I would certainly be interested to see more.

Sixteen years after that first Tucson conference, the science of consciousness has gone from strength to strength. But the state of play in the science of consciousness with respect to the hard problem is very much as it was. Almost everyone in the field recognizes that there is a hard problem of consciousness that poses special challenges (perhaps this is a sort of progress, although I think that most people recognized it all along). There have been many neurobiological and cognitive models of consciousness, but few of them have been offered as a solution to the hard problem, and when they have, hardly anyone has been convinced. Any associated claims about the hard problem are usually driven by bits of philosophy that are largely independent of the science and can be responded to straightforwardly. I have occasionally been invited to write critical reviews of this sort of work, but one bonus of the phraseology is that it generates a critique that writes itself: the work addresses the easy problems but not the hard problem, it provides a correlate of conscious experience but not an explanation, and so on. So I have not usually felt the need to go back over this ground.

That being said, positive nonreductive theories of consciousness have not had a much easier time of it. The speculative “theory” outlined in the second half of this chapter has certainly not gained widespread acceptance, and even I am inclined to be skeptical about the more speculative parts of it, such as the double-aspect theory of information. There have been a handful of attempts at nonreductive theories (the work of Stuart Hameroff and Gregg Rosenberg comes to mind), but not many, and none has attained much support. Perhaps it is too early in the science for positive theories that address the hard problem in a substantial and successful way. Still, I think it is worth trying.

In any case, the science of consciousness is thriving. Most of the work in the field does not try to solve the hard problem of consciousness but is none the worse for that. There are robust research programs in neurobiology (especially vision, but also emotion, bodily awareness, and other areas), psychology (especially conscious and unconscious processing in memory and learning, and the relationship between perception, attention, and consciousness), clinical neurology (especially on the vegetative state and other postcoma conditions), and other areas. These programs raise all sorts of fascinating philosophical issues, and in chapters 2–4 I address some of them.

Chapter 2 gives an overview of the science of consciousness and presents a picture of how it can work even in the absence of a solution to the hard problem. The central role of the science, as I see it, is to provide a bridge

between third-person data about brain and behavior and first-person data about consciousness, where it is data of the latter sort that make the science of consciousness truly distinctive. I suggest that the science is essentially correlative rather than reductive—at least, the central parts of the science make claims about correlation without making claims about reduction. Science of this sort can still have great explanatory power, however, by articulating powerful general principles that connect the first-person and third-person domains. The science of consciousness is not yet at the point of having that power, but if one squints, one can see the possibility of a framework of that sort in the distance. After discussing the role of much of the present work in the field, the chapter ends with discussion of some obstacles and of where the science might eventually go. I have tried to sketch a framework for the science that can be embraced by reductionists and nonreductionists alike.

I have given talks corresponding to chapter 2 at numerous conferences and other venues, including a memorable occasion at the Central Intelligence Agency in Virginia in 2000. (The talk was advertised without my knowledge as “Consciousness at the Millennium: The Mind-machine Connection”; I think the audience did not get what it was expecting, at least until I mentioned the consciousness meter.) Dan Dennett was present on a couple of occasions and responded in a debate at Northwestern with a paper titled “The Fantasy of First-person Science.” I have not responded to Dennett’s paper in print anywhere, so I have added an afterword indicating the general lines of my response, as well as discussing some other central issues involving first-person data.

Chapters 3 and 4 address what is arguably the centerpiece of the resurgence of the science of consciousness: the search for neural correlates of consciousness, especially of visual consciousness. Chapter 3 addresses conceptual issues about what it is to be a neural correlate of consciousness. The question might seem straightforward initially, but it generates all sorts of interesting puzzles. Must there be one neural correlate of consciousness, or might there be many? How strong a correlation is required? Over what range of cases? The chapter addresses these questions and a number of others and puts forward a proposal for understanding the notion of a neural correlate of consciousness in a way that does justice to the way the notion is used in the field. I use this analysis to draw some conclusions for the methodology of the search for neural correlates of consciousness.

Chapter 4 focuses directly on the epistemology of the search. The key question is: how can we isolate the neural correlate of consciousness in the absence of a “consciousness meter,” which measures consciousness directly? The obvious answer is that we rely on verbal reports and other behavioral

indicators of consciousness. Importantly, the use of these indicators tacitly relies on pre-experimental principles that connect these behavioral indicators to consciousness. The use of these principles is unavoidable, and it has a number of interesting consequences for the science. After drawing out some of the key principles, I once again draw some conclusions that bear directly on the practical methodology of the scientific work in this area.

Of course, there are many important issues in the science of consciousness that I do not address in these chapters. There is much to be said about the science of unconscious processes, especially regarding the philosophical assumptions behind the principles used in this area to ascribe conscious and unconscious mental states to subjects. I have discussed that work in talks (e.g., “Implicit Philosophy in Implicit Cognition Research”) but not yet in print. Likewise, there are very interesting philosophical issues concerning how to monitor consciousness in apparently unresponsive patients, such as those diagnosed with vegetative state; I have recently started some collaborative work with neurologists and neurobiologists on this topic. The relationship between consciousness and attention also raises all sorts of interesting issues that I hope to address in future work.

Chapters 5–7 focus on issues about materialism and dualism with regard to consciousness. The arguments of chapter 1 make a case for a view in which consciousness is irreducible and nonphysical, but much more needs to be said to flesh out these arguments, to answer objections, and to investigate the resulting views.

Chapter 5 provides an overview of the territory. It starts by presenting the central arguments against materialism, which involve establishing an epistemic gap between the physical and the phenomenal, and moves from there to an ontological gap. It then distinguishes between the three most important sorts of materialist opposition to these arguments: what I call type-A materialism (which denies the epistemic gap), type-B materialism (which accepts the epistemic gap but denies the ontological gap), and type-C materialism (which holds that there is a deep epistemic gap but one that will be closed in the limit). I make the case, in reasonable detail but without technicality, that each of these three views should be rejected. In the second half of the chapter I investigate the most important nonreductive views that result: type-D dualism (or interactionism), type-E dualism (or epiphenomenalism), and type-F monism (Russellian monism, or panprotopsychism). I discuss the pros and cons of each, suggesting that all three have significant attractions and that none has fatal flaws.

My own loyalties are fairly evenly spread among these three views, depending on the day of the week. I have often been told that I am an

epiphenomenalist, based on the sympathetic discussion of epiphenomenalism in *The Conscious Mind*. Sometimes that is taken to be sufficient reason to reject the arguments of the book, as epiphenomenalism is widely regarded as crazy. But even in *TCM I* I was just as sympathetic to Russellian monism (in both panpsychist and nonpanpsychist variants) as to epiphenomenalism, and Russellian monism involves a causal role for consciousness. In addition, I now think that the criticisms of interactionism there were much too quick (for reasons discussed in the section on type-D dualism here) and that there are no decisive reasons for rejecting that view. In any case, the proper conclusion of the anti-materialist arguments is disjunctive. The choice among the three disjuncts rests on further and largely independent considerations. As things stand, the choice is wide open. At the end of the day, the choice will come down to which of the disjuncts yields the most successful detailed theory, in light of a well-developed science of consciousness.

Chapter 6 is mainly devoted to the conceivability argument against materialism, viewed through the lens of the two-dimensional semantic framework. An argument of this sort was one of the centerpieces of *The Conscious Mind* and has attracted many replies. At the same time, the two-dimensional argument there was not as clearly formulated as it could have been. Here I have tried to give the argument a really clear formulation and to answer all of the central objections that have been raised against it. This chapter is unavoidably technical, but I would like to think that this is a case in which technicality has some rewards, not just in understanding issues about consciousness, but also in understanding the metaphysics and epistemology of modality.

The key issue in this chapter is whether conceivability (of some sort) entails metaphysical possibility (of some sort). The key opponent is the type-B materialist, who denies the entailment. Many objections and putative counterexamples to the conceivability–possibility thesis have been mooted: the chapter discusses fifteen or so putative counterexamples, along with ten or so objections of other sorts to the conceivability argument. I think that on close examination there are straightforward replies to most of them. Some raise deeper issues: about the apriority of modal epistemology, for example, or about the connection between concepts, rationality, and modality. Late in the chapter I sketch a positive grounding for the sort of modal rationalism that drives the conceivability argument. In an afterword I bring the analysis to bear on some other central arguments against materialism (the knowledge argument, Kripke's modal argument, the property dualism argument, the argument from disembodiment, the semantic stability argument), arguing that the issues underlying each are

closely related and that a two-dimensional analysis can shed useful light on each.

Chapter 7 (coauthored with Frank Jackson) addresses a related form of opposition to the anti-materialist arguments. Some type-B materialists allow that a unique epistemic gap exists between physics and consciousness: truths about consciousness are not deducible from physical truths, but truths about water, life, and other high-level phenomena are deducible from physical truths. Others argue that these epistemic gaps arise in many high-level domains. In a 1999 paper, Ned Block and Robert Stalnaker respond in the second way to arguments given by Jackson, by Joseph Levine, and by me, suggesting that truths about water, life, and heat are no more deducible from physical truths than are truths about consciousness. If so, the case of consciousness is nothing special, and unless one is prepared to accept that water is nonphysical, one should reject the argument that consciousness is nonphysical.

In response, Jackson and I argue that there are in fact a priori entailments from a nearly physical base to truths about water, life, and so on. The base needs to be expanded a little to allow indexicals, a “that’s all” truth, and of course truths about consciousness. But from this base, other ordinary macroscopic truths can be deduced by a priori reasoning. The argument here turns on some general observations about concepts and conceptual analysis. One of the key points is that there can be a priori entailments even in the absence of definitions or explicit analyses, contrary to what Block and Stalnaker appear to assume. The issues here are largely conceptual and epistemological, but we draw consequences for the explanatory gap regarding consciousness.

Chapters 6 and 7 raise issues in metaphysics, epistemology, and the philosophy of language that go well beyond issues about consciousness. Much more needs to be said about these issues, and in recent work I have tried to say some of it. A forthcoming book is devoted to issues in the philosophy of language and content, arguing for a Fregean approach to meaning and an internalist approach to mental content, grounded largely in the two-dimensional semantic framework that is central to chapter 6. Another forthcoming book takes up some of the issues in chapter 7 through the lens of the “scrutability” thesis, arguing that there is a limited class of truths from which all truths are deducible. I suggest that this scrutability thesis can be used to vindicate some (not all) of the aims of Rudolf Carnap and other logical empiricists, and that it has many interesting consequences in epistemology, the philosophy of language, metaphilosophy, and metaphysics. For me, these works grew out of seeds in work on consciousness and then took on a life of their own. This seems to me to

be a sign of the fertility of issues about consciousness for thinking about philosophy more generally, although perhaps it is just a sign of the way that every issue in philosophy is connected to every other.

Chapters 8 to 10 concern our thoughts and beliefs about consciousness, and the concepts we use to think about consciousness. These concepts are now known throughout philosophy as phenomenal concepts. Chapter 8 develops an account of the distinctive nature of phenomenal concepts, grounded in part in an analysis of the epistemological and conceptual observations that drive the arguments against materialism. I argue that phenomenal concepts behave in a way that is quite unlike most other concepts, involving a very strong sort of direct reference, on which the phenomenal qualities that are the referents of the concepts are also somehow present inside their sense. Here, the two-dimensional framework is again a useful tool in analyzing the phenomenon. The account in this chapter is in principle compatible with materialism, and in recent years some materialists have developed closely related accounts, but I think the account itself is fairly neutral (perhaps ultimately with some support for dualism, for reasons given toward the end of chapter 10). In any case, the issues concerning concepts and belief are of much interest in their own right.

Chapter 9 concerns our knowledge of consciousness. It was originally the second half of a long paper also containing chapter 8, and it builds on the central idea of that chapter. The special phenomenal concepts of chapter 8 lead to a distinctive class of “direct phenomenal beliefs,” which I argue have many interesting epistemological properties. For a start, they support a sort of infallibility thesis: direct phenomenal beliefs cannot be false. This thesis can do only limited epistemological work, but analysis of these beliefs leads to a more substantial epistemological view that involves a central role for acquaintance. I use the framework to analyze two important issues in the epistemology of consciousness: epistemological arguments against nonreductive views of consciousness, as well as Wilfrid Sellars’s arguments against the “given.” I suggest that with the appropriate analysis in hand, both sorts of arguments can be defanged. (Among other things, I think that this analysis provides a better response to the “paradox of phenomenal judgment” than I gave in *The Conscious Mind*.)

These chapters are explicitly concerned just with beliefs about consciousness, but much of the material here has application to a much broader class of beliefs. For example, I think that some perceptual beliefs are closely related to direct phenomenal beliefs, such that consciousness plays a central role in their constitution. Moreover, these consciousness-based beliefs may play a distinctive role in perceptual epistemology. The role for consciousness may

go further still. Chapter 8 concludes by drawing some speculative morals for the role of consciousness in grounding intentionality quite generally (a topic that I revisit in chapters 11 and 12). Chapter 9 ends with some morals about the general role of consciousness in epistemology.

Chapter 10 concerns a role for phenomenal concepts in the debate over materialism. As I noted earlier, many type-B materialists think that a unique epistemic gap exists between physical and phenomenal truths, one that does not arise in other domains. It is then incumbent on them to explain why there should be such a gap in a purely physical world. By far the most popular strategy is to argue that the gap results from the way we think about consciousness (rather than from the nature of consciousness itself) and in particular from the nature of phenomenal concepts. Many different accounts of phenomenal concepts have been offered for this purpose. In chapter 6 I argue against two such accounts individually. In chapter 10 I discuss the accounts as a class and argue that no such account can work. In particular, I argue that there is no account of phenomenal concepts such that the nature of the concepts is both explainable in physical terms and capable of explaining our epistemic situation regarding consciousness. And I argue that without doing these things, the account is toothless in explaining away the epistemic gap. Put differently, if an account of phenomenal concepts is substantial enough to explain our distinctive epistemic situation regarding consciousness, it will itself be as difficult to explain in physical terms as consciousness was. If I am right about this, it removes perhaps the most powerful sort of opposition to the arguments of chapters 1, 5, and 6.

Chapters 11–13 concern the contents of consciousness: in particular, the way that consciousness represents the external world. A central theme is the connection between consciousness and intentionality. Chapters 11 and 12 are concerned with the contents of perceptual experience. Chapter 13 is concerned with external-world representation in belief rather than in perception, but the themes are closely connected.

Chapter 11 focuses on representationalism, or intentionalism: roughly, the view that consciousness is essentially and wholly a matter of representing goings-on in the world. Representationalism has been a central view in recent philosophy of perception and philosophy of consciousness more generally. It is often put forward as a reductive thesis and is used in the interest of reducing consciousness to the physical. It is also often put forward as a radically externalist thesis, on which consciousness is grounded in states outside the head. These commitments are quite inessential to representationalism per se, however, and representationalism is much more plausible when it is detached from them. In this chapter I argue for a



distinctive sort of nonreductive, internalist representationalism, one that has most of the benefits of other sorts of representationalism along with relatively few of the costs.

The first half of chapter 11 surveys the territory. I clarify the nature of representationalism and argue for its plausibility. I survey varieties of representationalism, making the case for a nonreductive, internalist variety. The second half of the chapter develops an account of the contents of perceptual experience. If internalism is true, it appears that standard forms of environment-dependent content are ruled out and that one needs a sort of narrow (or internal) content instead. I bring in some ideas from the two-dimensional semantic framework to analyze the content of perception into two dimensions: Fregean content and Russellian content. Fregean content is narrow, while Russellian content is not. I argue that Fregean content is closely associated with the phenomenal character of experience. An afterword to the chapter spells out the application of the two-dimensional framework to perception in more depth.

Chapter 12 takes up the issue of the contents of experience where chapter 11 leaves off. I argue that the account of Fregean content in chapter 11 has some important inadequacies concerning its relation to phenomenology and leaves some crucial issues unexplained. I develop a further account involving what I call Edenic content: content that represents primitive properties (such as primitive redness or greenness), properties that are not instantiated in our world but that one can imagine might have been instantiated in the Garden of Eden. I bring this account to bear on many questions in the philosophy of perception, including questions about color constancy, spatial content, the representation of objects, and more.

I originally intended chapters 11 and 12 to be one long piece of work, but they are each long enough as they stand. Separating them might give the sense that I had a change of mind between the chapters, but that is not quite right: the Edenic view in chapter 12 extends the Fregean view in chapter 11 without rejecting it. When it comes to the metaphysics of consciousness, I think the Edenic story is more fundamental: fundamentally, consciousness may consist in the phenomenal representation of certain primitive properties. However, more than one sort of content is required for an adequate account of the many aspects of representation in experience. This reflects a general pluralism about mental and semantic content: one almost always needs contents of many different sorts to properly understand the complex phenomena of representation.

Chapters 11 and 12, along with chapters 8 and 9, can also be viewed as a contribution to what is now often called the “phenomenal intentionality” research program: the program of grounding intentionality in phenomenology.



Chapters 11 and 12 suggest that a distinctive sort of intentionality is inherent in perceptual experience. Chapters 8 and 9 suggest that phenomenology can play a constitutive role in determining the content of our beliefs, along with a corresponding epistemological role. These suggestions might well be taken much further. While I doubt that all of intentionality is fully determined by phenomenology, I think that intentionality cannot be properly understood without giving phenomenology a central role. I am at least attracted to a view in which all “narrow” intentionality is grounded in phenomenal intentionality along with functional and/or inferential roles and in which “wide” intentionality is grounded in turn in narrow intentionality plus the environment in the way suggested by the two-dimensional framework. I cannot claim to have made the case for a view as strong as that here, but I hope to develop it in future work.

Chapter 13 uses the movie *The Matrix* to address issues about our knowledge of the external world. On the face of it, the movie raises a version of Descartes’ skeptical challenge. Just as I cannot know I am not the victim of an evil genius, I also cannot know I am not in a matrix. And if I am in a matrix, so the challenge goes, most of my beliefs are false: I am not really seeing a table in front of me, I do not really live in Australia, and so on. In this chapter, I argue that this thought, although initially compelling, is wrong. Even if I am in a matrix, there are still tables and cars, and most of my beliefs remain true. That is, the hypothesis that I am in a matrix is not a skeptical hypothesis, as traditionally thought. Instead, it is a sort of metaphysical hypothesis about the underlying nature of our world. If we are in a matrix, the physical world is more fundamentally a computational world, in which things are made of bits. This is an interesting new metaphysics, but it does not lead to skepticism. I argue that much the same applies to many traditional skeptical hypotheses, such as the evil demon hypothesis and the hypothesis that my life is a dream. This does not provide a complete victory over skepticism (some skeptical hypotheses survive), but it nevertheless helps in the project of vindicating knowledge of the external world.

The conclusion of the chapter is in some ways reminiscent of antiskeptical conclusions by philosophers as diverse as Berkeley and Putnam. However, I arrive at the conclusion by a quite different route, one that involves reflection on the possible character of physics, as well as on the possible character of the mind-body relation and of creation. Most people start off very dubious about the conclusion (just as I did), but my experience is that the arguments bring a surprising number of people around. I am not sure what to make of this, except perhaps that in this domain people’s Cartesian intuitions run less deep than one might think.

On the face of it, chapter 13 might seem distant from chapters 11 and 12, with little explicit discussion of perception and consciousness. In fact the issues are deeply linked. If the arguments in this chapter are correct, the beliefs of a subject in a matrix are largely true. The same goes for perception: the perceptual experiences of a subject in a matrix are largely veridical. This picture coheres well with the Fregean picture developed in chapter 11, according to which perception represents properties that are the normal causes of certain sorts of perceptual experiences. This Fregean content does not require properties of a highly constrained sort, so if it turns out that that the relevant experiences are typically caused by certain computational properties, then these properties are represented by our experiences. This way, our experiences can be largely veridical, and our corresponding beliefs largely true.

Still, there is a persisting intuition that, if we are in a matrix, the world is not wholly as it seems. This intuition can be accommodated by bringing in ideas from chapter 12. If we are in a matrix, our experiences are not *perfectly veridical*: that would require our world to have primitive color properties, primitive spatiotemporal properties, and so on, as in Eden. But a moral of the Matrix chapter is that a matrix world is no better and no worse than a nonmatrix world in which quantum mechanics, relativity, and so on are true. In such a world, we have already fallen from Eden. Even in the world as revealed by contemporary science, it seems that our experiences are not perfectly veridical (their Edenic contents are not true): at best, they are imperfectly veridical (their ordinary Fregean and Russellian contents are true). But imperfect veridicality is good enough. And imperfect veridicality is something that the world of science shares with the world of the Matrix.

In effect, Eden and the Matrix can be seen as two poles in a familiar philosophical dichotomy. Eden is akin to the manifest image: the world exactly as it seems to us. The Matrix is (very loosely) akin to the scientific image: the messy and complicated real world of science with its many divergences from the manifest. But despite the mismatch between these images, we have grown used to the idea that the world of the scientific image is good enough to make our ordinary beliefs about the world true. The world does not vindicate them completely: a perfect match would require primitive solidity, primitive redness, and primitive squareness, and nothing in our world is really that way. So in a sense, our manifest image puts demands on the world that it does not meet. Still, the world of science vindicates our beliefs and experiences well enough by meeting the standard of imperfect veridicality. Perfect veridicality would require our world to be Eden, but imperfect veridicality can be satisfied even in a matrix.

These two poles call to mind another famous dichotomy. Eden corresponds at least loosely to Kant's phenomenal world: the world of things as they appear. The Matrix corresponds loosely to Kant's noumenal world: the world of things in themselves. We might think of Eden as a pure phenomenal world: if we were in Eden, nothing would have a hidden nature. But if we are in the Matrix, the world has a hidden nature that is not revealed to us in perception or even in ordinary science. Like the analogy with the scientific image, the analogy with the noumenal world is imperfect. The movie offers us an un-Kantian route to see the things in themselves: the red pill reveals their computational nature, which of course just raises the question of the noumena underlying the computers in turn. Still, through this lens, the film no longer seems to be an illustration of Cartesian deception. Rather, it is an illustration of Kantian (or perhaps Russellian) humility. Perception and even science do not reveal the entire intrinsic character of the world. Perhaps it was unreasonable ever to expect that they would. Still, even if perception and science are not perfect, in their imperfect ways they are adequate to their tasks all the same.

Reflection on the Matrix scenario is fertile in many other ways. Coming to grips with it requires getting to the bottom of some of the deepest issues in philosophy. These include not just issues in epistemology but also those in metaphysics (what is the nature of objects?), the philosophy of language (what determines the reference of our expressions?), the philosophy of mind (how do we make cognitive contact with the world?), the philosophy of perception (how do we represent space?), the philosophy of physics (what is the role of spatiotemporal concepts in theoretical physics?), the philosophy of computation (what does it take to implement a computation?), and even the philosophy of religion (what should we say about God if our world is the product of imperfect creators?) and ethics (is life in a matrix as meaningful as a life outside it?). I address a few of these issues in the chapter and in the philosophical notes that follow. I am sure that I have not gotten to the bottom of any of them. Still, because of the way that it opens a door to all of these issues in such a simple way, this chapter is my favorite in the book.

Chapter 14 (coauthored with Tim Bayne) addresses another Kantian topic: the unity of consciousness. It is hard to turn this fuzzy topic into clear philosophy, but in recent years a few people have tried. This is our attempt. We start with the question of what it is for two states of consciousness (experiencing red and hearing middle C, say) to be unified. We analyze a number of different notions of unity in search of one that undergirds a nontrivial but still plausible unity thesis holding that consciousness is necessarily unified. This leads to a notion of phenomenal unity, defined

in terms of a sort of conjoint phenomenal character. We suggest that there is a *prima facie* case that consciousness is necessarily unified in this sense and that *prima facie* counterevidence from split-brain syndrome can be accommodated by making relevant distinctions. To get a better handle on unity, we introduce a quasi-mereological analysis and an analysis in terms of entailment between phenomenal states. We use these analyses to argue against certain representationalist and higher-order theories of consciousness on the grounds that they cannot vindicate the unity thesis. We end by speculating about why consciousness is unified, if indeed it is.

The end of the chapter suggests a certain holistic view of the metaphysics of consciousness, one in which local states of consciousness (experiencing red, hearing middle C) are grounded in total states of consciousness for a subject at a time. It is tempting to think that these states are the basic units of consciousness. If we combine this picture with the picture in chapter 12, a more detailed hypothesis in the metaphysics of consciousness suggests itself: consciousness consists in the unified phenomenal representation of an Edenic world. I do not know whether this hypothesis is correct, but it is at least aesthetically pleasing.

The appendix provides an introduction to the two-dimensional semantic framework that I use in a number of the more technical chapters. This framework provides a very useful tool for analyzing the relationship between modal and epistemological issues, as well as for modeling the content of linguistic expressions and mental states in a way that is sensitive to their cognitive and epistemological roles. After outlining the framework and some of its applications, I address a number of objections that various philosophers have put forward. I think that once the framework is understood correctly, there are straightforward replies to many of these objections, although of course some objections raise substantial issues, too. But readers can decide for themselves.