HOUSE OF MIRRORS

The Illusion of Phenomenal Consciousness

François Kammerer

OXFORD

House of Mirrors

PHILOSOPHY OF MIND SERIES

SERIES EDITOR: David J. Chalmers, *New York University*

The Conscious Brain
Jesse Prinz

Simulating Minds
The Philosophy, Psychology, and Neuroscience of Mindreading
Alvin I. Goldman

Supersizing The Mind
Embodiment, Action, and Cognitive Extension
Andy Clark

Perception, Hallucination, and Illusion
William Fish

Phenomenal Concepts and Phenomenal Knowledge
New Essays on Consciousness and Physicalism
Torin Alter and Sven Walter

The Character of Consciousness
David J. Chalmers

The Senses
Classic and Contemporary Philosophical Perspectives
Fiona Macpherson

Attention Is Cognitive Unison
An Essay in Philosophical Psychology
Christopher Mole

The Contents of Visual Experience
Susanna Siegel

Consciousness and the Prospects of Physicalism
Derk Pereboom

Consciousness and Fundamental Reality
Philip Goff

The Phenomenal Basis of Intentionality
Angela Mendelovici

Seeing and Saying
The Language of Perception and the Representational View of Experience
Berit Brogaard

Perceptual Learning
The Flexibility of the Senses
Kevin Connolly

Combining Minds
How to Think about Composite Subjectivity
Luke Roelofs

The Epistemic Role of Consciousness
Declan Smithies

The Epistemology of Non-Visual Perception
Berit Brogaard and Dimitria Electra Gatzia

What Are Mental Representations?
Edited by Joulia Smortchkova, Krzysztof Dołęga, and Tobias Schlicht

Phenomenal Intentionality
George Graham, John Tienson, and Terry Horgan

Feminist Philosophy of Mind
Keya Maitra and Jennifer McWeeny

The Border Between Seeing and Thinking
Ned Block

The View from Everywhere
Realist Idealism without God
Helen Yetter-Chappell

House of Mirrors
The Illusion of Phenomenal Consciousness
François Kammerer

House of Mirrors

The Illusion of Phenomenal Consciousness

FRANÇOIS KAMMERER

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries.

Published in the United States of America by Oxford University Press
546 Fifth Avenue, New York, NY 10036, United States of America.

Library of Congress Cataloging-in-Publication Data
Names: Kammerer, François author
Title: House of mirrors : the illusion of phenomenal consciousness /
François Kammerer, CNRS, Archives Henri-Poincaré.
Description: New York, NY : Oxford University Press, [2026] |
Includes bibliographical references and index.
Identifiers: LCCN 2026008204 (print) | LCCN 2026008205 (ebook) |
ISBN 9780197840252 hardback | ISBN 9780197840283 | ISBN 9780197840269 epub
Subjects: LCSH: Consciousness | Phenomenology
Classification: LCC B808.9 .K36 2026 (print) | LCC B808.9 (ebook)
LC record available at https://lccn.loc.gov/2026008204
LC ebook record available at https://lccn.loc.gov/2026008205

DOI: 10.1093/9780197840283.001.0001

Printed by Integrated Books International, United States of America

The manufacturer's authorized representative in the EU for product safety is
Oxford University Press España S.A. of Parque Empresarial San Fernando de Henares,
Avenida de Castilla, 2 – 28830 Madrid (www.oup.es/en or product.safety@oup.com).
OUP España S.A. also acts as importer into Spain of products made by the manufacturer.

For Ismahan

Preface

When a novel philosophical idea first comes to mind, the moment is delightful. The idea is beautiful, sparkling, intoxicating. Then we reflect, we read. We understand that we must explore it further. Years of work follow, and all that labor transforms the initial idea. We realize that it is not so novel after all. We see that it might not solve the world's mysteries, but only a local problem. And then, properly formulated, defined, argued, distinguished from its counterparts, the idea gains in clarity and modesty, at the price of becoming what feels like a banality—at best, we still dare to hope, a true banality.

This book is the result of years of work. It presents a set of ideas about consciousness that I started developing around 2013, that I once found extraordinarily exciting, and that now tend to appear to me, for the most part, and when I approach them in an academic context, as evident truths. The hope that they may not be obvious to everyone, that others may find them informative, enlightening, interesting even, is the motivation behind this book. The first joys an idea brings are the solitary and confident joys of discovery. The last are anxious and social joys: joys by proxy.

* * *

My thinking about consciousness owes much to many people. I would not have written this book in this manner if it were not for various conversations, real or imagined, with countless other thinkers, dead or alive. Focusing on those I have met at least once and who shaped the most my understanding of consciousness and my development as a philosopher, whether or not they realized it, I want to thank, in approximate chronological order of encounter: Dominique Desenne, Pascal Ludwig, Daniel Andler, Joseph Levine, Derk Pereboom, David Chalmers, Uriah Kriegel, David Papineau, Keith Frankish, Michael Graziano, Eric Schwitzgebel, Tobias Schlicht, Matthias Michel, Anna Giustina, Andrew Lee, Daniel Dennett, Hedda Hassel Mørch, Nicholas Humphrey, and Jonathan Birch.

Some people were kind enough to provide me with detailed comments on earlier versions of this book: Christopher Devlin Brown, Leonard Dung,

Keith Frankish, Keith Harris, François Jaquet, Andrew Lee, Matthias Michel, Hedda Hassel Mørch, Luke Roelofs, Eric Schwitzgebel, and Christian de Weerd. I want to thank them all very much for their incredibly valuable feedback.

Many people and institutions have actively supported me in recent years. I would not have managed to write this book without the support of Tobias Schlicht, David Chalmers, the Belgian FNRS, the Humboldt Stiftung, the Ruhr-Universität Bochum, New York University, and, more recently, the French CNRS. I particularly want to thank the CNRS for making the open access publication of this book possible.

Finally, on top of the people I already mentioned, many persons have, through their questions, their conversations, or their writings, contributed, in some way or other, to the work leading to this book. This list cannot be exhaustive, but I would like to thank Katalin Balog, Arianna Beghetto, Alexandre Billon, Ned Block, Romain Bourdoncle, Richard Brown, Heather Browning, Peter Carruthers, Tim Crane, Nicolas Delon, Erhan Demircioğlu, Krys Dolega, Marco Facchin, Katalin Farkas, Oscar Gilg, Thibaut Giraud, Philip Goff, Liz Irvine, Max Kistler, Michelle Liu, Florian Marion, Jakub Mihálik, Fabien Mikol, Takuya Niikawa, David Rosenthal, Rob Rupert, Dan Ryder, Maja Spener, Tobias Starzak, Daniel Stoljar, Assaf Weksler, Wanja Wiese, as well as colleagues at the Sorbonne, the Institut Jean Nicod, the UCLouvain, the Ruhr-Universität Bochum, New York University, the CNRS, and the Archives Henri-Poincaré.

I want to thank Peter Ohlin, Meredith Taylor, Cameryn Cole, and Leo Mosquline for their editorial work.

I want to thank my family and my friends for their support throughout these years and for not looking too surprised when I told them that, yes, I am still writing this book.

Finally, and above everyone else, I want to thank Ismahan, Sarah, and Elias. Writing a book that aims at truth constitutes, in its own strange way, an act of love, which was only made possible and meaningful thanks to the very different—and infinitely more significant—sort of love they bring to my life; not to mention Ismahan's supernatural patience, kindness, and generosity, which amaze me every day.

Contents

Introduction

This book is about phenomenal consciousness. It has three goals.

Its first goal is to present and defend a view I think is true, but counterintuitive. It is a variety of *illusionism*. According to illusionism, phenomenal consciousness does not exist, although it seems to exist. Unlike some other versions of illusionism, my view emphasizes that the illusion of phenomenal consciousness is *not* the result of faulty theorizing or sloppy reasoning. It is not a philosopher's dream, nor a theorist's mistake. Our belief in phenomenal consciousness is false, but it is coherent and highly intuitive, even when we reason impeccably. The illusion of phenomenal consciousness is generated by robust features of our introspection. It can be uncovered, but never entirely dispelled. This illusion of phenomenal consciousness is even stronger than other robust illusions, like perceptual illusions. The very idea that phenomenal consciousness is illusory is extraordinarily hard to grasp, which makes the illusion of phenomenal consciousness uniquely strong and almost inextricable.

The second goal of this book is to explain *why* the illusion of phenomenal consciousness arises, and why it is uniquely strong and almost inextricable. To do so, I develop a view I call Meta-Cartesianism. In this view, when we introspect phenomenal experiences, we represent what appears to us the most immediately—the most direct evidence we possess. These states of direct evidence are grasped as presenting other things to us. They are also grasped as *themselves* presented to us in a uniquely infallible and revelatory manner. However, these introspective representations are misrepresentations. We do not really enter mental states with this peculiar presentational profile.

My view is Meta-Cartesian, but not Cartesian. It says that Descartes was incorrect about what our minds truly are: there are no phenomenal experiences. Still, his conception was a relatively accurate transcript of what our minds introspectively appear to be.

Meta-Cartesianism views our representations of phenomenal experiences as closely intertwined with our *evidential* cognition—our intuitive way of reasoning about evidence, appearance, and illusions. This tight entanglement

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0001

is what generates deep and unique conceptual difficulties when we try to apprehend phenomenal consciousness as an illusion. These difficulties do not arise when we think of other illusions. They are what make the illusion of phenomenal consciousness uniquely strong and almost inextricable. However, if we correctly understand this entanglement and use the right concepts, we can coherently apprehend the illusory nature of phenomenal consciousness, and appreciate the force of illusionism. The illusion of phenomenal consciousness is a house of mirrors. To escape it, we must close our eyes and let the handrail be our guide.

This book has a third goal. It maps the implications of illusionism in various domains. First, consciousness science. If phenomenal consciousness is an illusion, what happens to the scientific study of consciousness? Should it be abandoned? Should it continue essentially untouched? Should it be reformed, and if so, how? Second, ethics. Phenomenal consciousness is often thought to play an essential ethical role. Many think that it is crucial for well-being or moral status. Should we change our ethical views once we embrace illusionism, and if so, how? Third, epistemology. Phenomenal consciousness is frequently seen as central to epistemic justification. How should our view of justification change once we accept illusionism?

I will show that, in these three domains, illusionism has non-trivial but also non-simplistic implications. These implications must be taken into account if, as I hope I will demonstrate, illusionism has a reasonable chance of being true.

* * *

The book comprises ten chapters, which can be roughly put in three groups, each corresponding to one of the three goals of the book.

In the first three chapters (Chapters 1–3), I pursue the first goal of the book: I present a version of illusionism, and I argue for it. In Chapter 1 ("A Bundle of Mysteries: Approaching Phenomenal Consciousness"), I define phenomenal consciousness. I explain that it seems *mysterious*, in the sense that it compellingly appears to us to have various mysterious peculiarities. In Chapter 2 ("Illusionism About Phenomenal Consciousness"), I present illusionism: the view that phenomenal consciousness does not exist, but simply seems to exist. I review versions of the view and zero in on my own. I then give a brief argument for the relatively uncontroversial claim that phenomenal consciousness seems to exist, and I lay the foundations for the argument in favor of the highly controversial claim that it does not

exist. In Chapter 3 ("Arguing for Illusionism"), I argue for this claim, with two arguments working in tandem: the *Argument from Mysteries* and the *Argument from Irrelevance*.

In the following three chapters (Chapters 4–6), I pursue the second goal of the book: I explain why the illusion of phenomenal consciousness arises and why it is uniquely strong. In Chapter 4 ("Approaching the House of Mirrors: The Illusion Problem"), I present the illusion problem: the problem of explaining the illusion of phenomenal consciousness. I critically examine extant candidate solutions, showing that they do not explain the unique strength of the illusion of phenomenal consciousness. In Chapter 5 ("Mapping the House of Mirrors: Introducing Meta-Cartesianism"), I present Meta-Cartesianism. For Meta-Cartesianism, phenomenal cognition and evidential cognition are deeply entangled. We represent phenomenal experiences *as* mental states endowed with certain evidential peculiarities. In Chapter 6 ("Leaving the House of Mirrors: Solving the Illusion Problem"), I show how Meta-Cartesianism explains crucial aspects of the illusion of phenomenal consciousness and solves the illusion problem.

Chapter 7 ("Addressing Objections") serves as a complement to the first three chapters of the book. It defends illusionism against various objections.

In the last three chapters (Chapters 8–10), I pursue the third and last goal of the book: I map the implications of illusionism in various domains. In Chapter 8 ("The Science of Sirens: Illusionism and Consciousness Science"), I examine illusionism's implications for the science of consciousness. In Chapter 9 ("Losing Your Religion: Ethics After Illusionism"), I explore its ethical implications, focusing on well-being and moral status. In Chapter 10 ("Shedding Our Cartesian Scales: Illusionism and Epistemology"), I examine its epistemological implications, focusing on epistemic justification.

Finally, in a concluding section ("Coda: Consciousness, Illusion, and Reality"), I briefly speculate on what illusionism suggests regarding the ultimate nature of reality.

1

A Bundle of Mysteries

Approaching Phenomenal Consciousness

How a material, physical process [. . .] could become a representation, an act of consciousness, is completely incomprehensible. [This] riddle will probably remain unsolved until the end of time; and I believe that, if today an angel came down from heaven and explained it all to us, our intelligence would not even be able to comprehend it.

—Wilhelm Griesinger

1.1. Phenomenal consciousness

This book is about *phenomenal consciousness*. What do I mean by this? Put yourself in the right mindset. Focus on how things feel to you right now. You are *seeing* a white page with inked letters and *smelling* coffee, while *recalling* the taste of the croissant you just finished and *imagining* the face of your dietitian learning about your indulgence.

Seeing, smelling, reminiscing, imagining: those are examples of what philosophers call "phenomenal experiences"—or "phenomenal states," "phenomenally conscious states," "raw feels," "qualitative states," etc. These states all seem to *feel* a certain way. *Phenomenal consciousness* is the overarching phenomenon. Phenomenal properties (or "qualitative properties," "qualia," etc.) are properties of mental states in virtue of which they are phenomenal experiences of certain types. Phenomenal whiteness, for instance, is a property of certain visual phenomenal experiences had when looking at white things. Taken together, the phenomenal properties of an experience fix its *phenomenal character*: the way it phenomenally feels.

Phenomenally conscious beings are beings with phenomenal experiences. They can also be called "*sentient*": *sentience*, in a broad sense, is the capacity to have phenomenal experiences. In a narrow sense, it is the capacity to have specifically *valenced* phenomenal experiences, which are attractive or

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0002

aversive—like pain and pleasure, joy and sorrow. Unqualified, "sentience" will here mean narrow sentience.

In what follows, I argue that phenomenal experiences do not exist, but only appear to exist. For now, I write as if they are really there, as they appear to you.

Many things happen to you that are not phenomenal experiences. Your autonomic nervous system regulates your heartbeat. Your digestive system deals with the croissant you ate earlier. Your immune system fights the virus transmitted yesterday by a coughing colleague. These things happen, but they do not *feel* like anything. Metaphorically, they take place "in the dark." There is *something it is like* for you to see white, or smell coffee. There is nothing it is like for you to have your autonomic nervous system regulate your heartbeat.

So, what are phenomenal experiences? Here is a first answer: they are mental states such that there is *something it is like* to be in them.[1] However, only seasoned philosophers find this definition satisfying. First-year students do not buy it, nor does the proverbial person on the street. "What it is like"—the locution is vague, metaphorical, and translates poorly in many languages.

Many admit that we cannot define "phenomenal consciousness" further, but think that this is not a problem. We know what phenomenal consciousness is. It is the most obvious phenomenon there is. It is present wherever we look—the fabric of our mental lives. The "what it is like" locution is not a full-blown definition but a tool—a definitional prop. With the right examples (seeing, smelling, imagining, etc.), it makes us pay attention to the phenomenon, which is so obvious that we might otherwise miss it, as we can forget about the air we breathe—as a fish can forget about water. Borrowing the answer Louis Armstrong gave when asked to define *"jazz"*, philosopher Ned Block claimed, "only half in jest," that "if you have to ask what consciousness is, you'll never know."[2]

This is a standard way of defining "phenomenal consciousness." It involves directing our attention to the right aspects of specific paradigmatic mental states, with the help of locutions such as "what it is like."[3]

As I said, I think phenomenal consciousness does not exist. Still, there is something legitimate in this definitional approach. Appropriately focusing our attention makes us grasp what phenomenal consciousness is—or rather,

[1] Thomas Nagel (1974) popularized the expression.
[2] Block (1978).
[3] For instance, see Searle (1992, 83); Block (1995, 230–31); Chalmers (1996, 4); Güzeldere (1997, 6–7); Schwitzgebel (2016b); Seth and Bayne (2022, 440).

what it *would be* if it existed. It allows us to think and speak meaningfully about it.

1.2. Phenomenal concepts

When we pay attention to how things feel to us in this way, we engage in introspection—more precisely, in *phenomenal* introspection. We thus form and deploy concepts of how our current mental episodes feel, of their phenomenal properties. These concepts are *phenomenal concepts*.

With phenomenal concepts, we can think thoughts expressed by sentences like "*this* is different from what I felt yesterday at the same hour"; "*this* feels good"; "I wonder if my friend sitting next to me also feels *this* right now." Here, I distinguish between *concepts*—constituents of thoughts—and *terms*—constituents of sentences, which can express concepts. I indicate terms and sentences with quotation marks, and concepts and thoughts with small caps: "this feels good" expresses THIS FEELS GOOD.

There are various ways to approach phenomenal introspection. It might be a non-conceptual representation of our experiences, which later enables their conceptualization. Alternatively, it might be a conceptual process through and through—the formation and deployment of phenomenal concepts. I presuppose the second view in what follows, but much of what I say could be preserved, with reformulations, if phenomenal introspection is non-conceptual.

Phenomenal introspection mobilizes phenomenal concepts, but phenomenal concepts are also deployed *outside* of introspection—for instance, to think about past or imagined phenomenal experiences (YESTERDAY I FELT THIS; SEEING AN AURORA BOREALIS PROBABLY FEELS LIKE THIS). Some phenomenal concepts are *composed* of other phenomenal concepts, like the concept of an experience of a red triangle on a white background. Others are basic: maybe, for instance, our concept of phenomenal whiteness.

Phenomenal concepts provide our grasp of what phenomenal experiences are—or would be, if they existed. I assume that they are *conceptually independent* from physical and functional concepts.[4] On a first approximation, this means that, from propositions about physical or functional properties, we cannot deduce a priori, using only logical and conceptual truths,

[4] Loar (1997).

non-trivial propositions about phenomenal properties. From propositions about phenomenal properties, we cannot deduce a priori, using only logical and conceptual truths, non-trivial propositions about physical or functional properties.

Why believe in this independence? Because dissociations between the physico-functional and the phenomenal are conceivable. We can conceive of *zombies*: creatures physically and functionally identical to phenomenally conscious humans, but deprived of phenomenal experiences.[5] We can conceive of cases of "fading" and "dancing" qualia: subjects whose phenomenal experiences change without physical or functional changes.[6] We can conceive of cases of phenomenal *inversions*: subjects who are physically and functionally identical but have inverted color experiences. We can conceive of 'floating' phenomenal minds: subjects with phenomenal experiences, but with no physical properties, and whose phenomenal experiences do not play any causal roles. All these decorrelations are conceivable. They are *negatively* conceivable: they can be thought of without contradiction. They are also, I think, *positively* conceivable: we can form a positive and coherent conception of them.[7]

Not everyone thinks that phenomenal concepts are independent from physical and functional concepts. Take analytic functionalists or logical behaviorists, for whom all mental concepts, including phenomenal concepts, are analyzable into functional or behavioral concepts.[8] They deny phenomenal concepts' independence. Defenders of the "phenomenal powers" view, for whom phenomenal concepts have a priori links to causal concepts, also deny it.[9] Still, these exceptions set aside, phenomenal concepts are widely assumed to be independent from physical and functional concepts.

Here is another assumption I make. Phenomenal concepts provide a substantive grasp of their referents, which they positively characterize. For instance, our concept of phenomenal whiteness gives us an idea of what phenomenal whiteness is. This idea might be undefinable in terms of other concepts, and uncommunicable to someone who does not have the concept.

[5] Kirk (1974).
[6] Chalmers (1996).
[7] For more, see Chalmers (2002b). Chalmers also distinguishes between *prima facie* and *ideal* conceivability. I take the decorrelations mentioned above to be ideally conceivable. Chalmers also understands "negatively conceivable" slightly differently, as meaning that a proposition cannot be ruled out a priori.
[8] These views are a variety of what has been called "type-A materialism" (Chalmers 2002a).
[9] Mørch (2020). See my discussion in Chapter 3.

Still, it is a substantive and positive idea. It allows us to reflect and see how phenomenal whiteness differs from phenomenal grayness and blackness, from a phenomenal smell of coffee, or a phenomenal pain in the shoulder. It also allows us to see what all these have in common but do not share with tables, churches, pizzas, numbers, colors (as properties of surfaces), stars, electrons, jokes, equations, and revolutions: by abstraction, we can form a positive and substantive *general* concept of *phenomenality*.

Some deny that phenomenal concepts provide a substantive grasp of their referents. In some views, phenomenal concepts are cognitively empty. They refer directly, without characterizing their referents.[10] The concepts' emptiness supposedly explains their independence: being cognitively empty, phenomenal concepts cannot be a priori 'hooked' to other concepts. This view of phenomenal concepts has been heavily criticized—successfully, in my opinion.[11] It is much more plausible that phenomenal concepts provide a substantive grasp of their referent. This explains why we can think and talk meaningfully about the phenomenal even if we cannot define it except with examples and vague locutions.

Conceptually, phenomenal consciousness differs from other things we call "consciousness," like access-consciousness or self-consciousness.[12] A mental state is *access-conscious* when its content is poised for use in reasoning and rational control of speech and action. A creature possesses *self-consciousness* when it can deploy a concept of oneself as oneself—e.g., when it can recognize itself in a mirror. Whether phenomenal consciousness is *really* distinct from these is highly debated. Here, I assume that the *conceptual* distinction holds.

Why accept this conceptual distinction? First, it is plausible that access-consciousness and self-consciousness are functionally defined. Therefore, phenomenal concepts' independence vis-à-vis functional concepts also renders them independent from concepts of access-consciousness and self-consciousness. Second, the distinction is directly intuitive. If phenomenal and access-consciousness were not conceptually distinct, we could not conceive of one without the other, but it seems that we can. As Ned Block pointed out, we can imagine being engaged in a lively conversation, without being access-conscious of the sound of the pneumatic drill outside, while

[10] See Carruthers (2000); Levin (2007); and Tye (2003). These views follow the "Phenomenal Concept Strategy" (Balog 2009; Stoljar 2005), associated with type-B materialism (Chalmers 2002a).
[11] Levine (2007); Demircioğlu (2013); and Tartaglia (2013).
[12] Block (1995).

phenomenally experiencing this sound. We can imagine a robot sufficiently functionally similar to us as to be access-conscious and self-conscious, but without phenomenal experiences.[13]

Let us be careful. By pointing out that the concept of phenomenal consciousness is distinct from concepts of access-consciousness or self-consciousness, or physical and functional concepts, we do not construct, by stipulation, a concept of phenomenal consciousness distinct from these other concepts. Things work the other way around. We already have phenomenal concepts, which we deploy, for instance, in introspection. These phenomenal concepts happen to be distinct from the other concepts I mentioned. This distinction then serves as yet another prop to make sure we associate the expression "phenomenal consciousness" with the right concepts we already possess.

Earlier, I distinguished *concepts*, which feature in thoughts, and *terms*, which feature in sentences. "Phenomenal consciousness" is a term—a term of art. It expresses a general concept of phenomenality, formed by abstraction, from individual phenomenal concepts. "Consciousness," however, is not a term of art. It is a common term, used in everyday talk. What does "consciousness" (tout court) mean? Which concepts do we express when we say that someone is conscious or has a conscious pain?

Following David Papineau,[14] I take many ordinary uses of "consciousness" (and cognate terms) to be *equivocal*. They might express phenomenal concepts, but also other non-phenomenal concepts of consciousness, which grasp their referents in terms of functional roles. Among the relevant non-phenomenal concepts, one finds notably, though not exclusively, concepts in the vicinity of those mentioned above: concepts of access-consciousness, self-consciousness, or rougher pre-theoretical versions of those. Papineau suggests that, in the typical case, ordinary uses of "consciousness" express a complex multi-criterial concept of consciousness. Phenomenal concepts only correspond to one of those criteria, while other criteria correspond to these other concepts just mentioned. Given this view, ordinary talk of *conscious pain* typically expresses a multi-criterial concept, plausibly satisfied by something with the right phenomenal property—a *phenomenally conscious* pain—but also, maybe, by something with the right functional property—e.g., an *access-conscious pain*. In everyday contexts, this is unproblematic, as

[13] Block (1995, 234).
[14] Papineau (2002, 97–103).

we usually assume that the properties corresponding to these criteria *always come together*. Distinguishing them requires reflection. It only makes sense in theoretical contexts.

So, we often deploy phenomenal concepts, but we rarely deploy them alone. In ordinary contexts, we often deploy them together with other concepts, to represent different criteria of multi-criterial concepts. In everyday talk, we rarely make utterances that unambiguously express only phenomenal concepts, and we have no terms to do so. Conceptual analysis is needed to introduce a dedicated technical term, "phenomenal consciousness," which exclusively expresses the concept of phenomenal consciousness.

1.3. The intuition of obviousness

When it comes to phenomenal consciousness, things are not as they seem. To understand what this implies, we must first explore the realm of appearances. What does phenomenal consciousness *appear* to be?

Some preliminaries are needed. I use "appear" and "seem" interchangeably. Those are polysemic. For instance, they have a *phenomenal* and an *epistemic* sense.[15] Read *epistemically*, "A seems F to me" means that I have some reason (or justification) to believe that A is F. Read *phenomenally*, it means that I have a phenomenal experience of A being F. Intuitively, both types of seemings are related. Phenomenal seemings ground epistemic seemings: experiencing a blue table gives me a reason to believe there is a blue table. I will later explain why this is intuitive. However, what matters here is the conceptual *distinction* between seemings. Moreover, "appear" and "seem" also have a third *psychological*—or *functional*—sense. In this sense, "A seems F to me" means that I am disposed to believe that A is F. Psychological seemings, too, might be related to other seemings, but again, what matters here is the conceptual distinction.[16]

So, what does phenomenal consciousness appear to be? For now, take the question to be about *psychological* appearances. Here is a response. Phenomenal consciousness appears mysterious. We have *mystery intuitions* about the phenomenal. "Intuition" refers here to psychological appearances with appropriately abstract content.

[15] See Jackson's analysis of looks (1977, 30–33). See also Chisholm (1957) and M. Liu (2024).

[16] We can distinguish *prima facie* vs. *ultima facie* epistemic appearances, and *subjective* (e.g., A appears F *to me*) vs. *objective* (e.g., iron pyrite has the appearance of gold) ones. I focus on subjective appearances, and prima facie epistemic appearances.

The first apparent mystery of phenomenal consciousness is that it seems *introspectively obvious*. We have an intuition of obviousness. On a first approach, this means, to put it metaphorically, that we are disposed to believe that our phenomenal experiences are immediately present under our mind's eye, standing in some uniquely intimate relation with us.

This apparent introspective obviousness does not concern all properties of our experiences. Take their extrinsic and relational properties. My pain experience happens on a Thursday, but this does not seem introspectively obvious. Take their causal properties. My pain experience disposes me to be rude with my friends, but this does not seem introspectively obvious. What seems obvious are only the *phenomenal properties* of experiences: what it is like to have these experiences. Intuitively, these properties are *intrinsic*. They concern experiences themselves, independently of their relations to other things. They are also *categorical*, not causal or dispositional. They regard what experiences are, not what they do.

Going deeper, this obviousness has two aspects. The first is *self-revelation*. Phenomenal properties of experiences seem introspectively presented, to the subject who has them, with their nature—the *nature* or *essence* of a property being what makes it what it is.[17] For instance, when I feel pain, the *phenomenal features* of pain, and what makes these features what they are, seem presented to me. I struggle to see how someone could phenomenally experience pain without being somehow presented with the aspect of pain that makes it phenomenally painful. If so, the phenomenal is peculiar. Other things can exist without being presented. I can have a bill in my wallet, a pimple on my shoulder, and low iron in my blood, without any of this being presented to anyone. When the non-phenomenal is presented, it is not usually with its essence. Looking at a lake, water is visually presented to me, but not with its essence of being H_2O.

This obviousness has a second aspect: the *infallibility* of introspective appearances. It seems that, on the basis of introspection, certain errors cannot be made about phenomenal experiences: errors based on *misleading presentations*.[18] It seems that, if I am introspectively presented with an experience having phenomenal property P, I must really have an experience

[17] Following M. Liu (2024, 3), I use an Aristotelian/Finean view of essence, on which the essence of something is what makes the thing what it is—what Fine (1995) calls the "immediate constitutive essence."

[18] Which have been called "errors of ignorance," as opposed to "errors of judgment" (Hill 1991, 127–28), or "brute errors," as opposed to "basing errors" (Smithies 2013, 1180).

with P. Intuitively, I can make errors about the phenomenal properties of my experiences, but only if I fail to record how things introspectively appear to me. So, my introspective *judgments* about my phenomenal properties can be wrong, but intuitively, introspection itself cannot misleadingly present my phenomenal properties. This also makes the phenomenal peculiar. When it comes to the non-phenomenal, errors based on misleading appearances seem possible. It can perceptually appear to me that I face a green pen, even if there is just a white pen (visual illusion), or no pen at all (hallucination). It can seem to me that I am a French philosopher named François, even if I am not, if I am the victim of a Truman-Show-like conspiracy.

Let us summarize. We intuit phenomenal consciousness to be introspectively obvious. It seems *self-revealing*: it seems that, if a subject has a phenomenal experience, the nature of the phenomenal properties of this experience introspectively appears to the subject. It seems *infallibly introspectively appearing*: it seems that, if it introspectively appears to a subject that they have an experience with phenomenal property P, they really have an experience with P.

I described the content of this intuition in sophisticated terms. Is it really plausible, then, that many of us have it? After all, almost no one in the world has an opinion on the revelation of phenomenal essences.

Still, one can have the right intuition without entertaining these beliefs. To have the intuition, it is enough if, given sufficient reflection and mastery of other relevant concepts, one has *some* (resistible) tendency to have these beliefs. More on this later.

The intuition of obviousness can be approached less technically. Earlier, I gave a metaphorical characterization: it is the intuition that our phenomenal experiences are immediately present under our mind's eye, standing in some uniquely intimate relation with us. Other non-technical characterizations are possible. The intuition of obviousness is the intuition that the *introspective appearance* of phenomenal properties is indistinguishable from their reality. It is the intuition that phenomenal features are presented to the subject through a sort of mental contact, in which one is "somehow 'so close' to the qualities of [their] experience that their nature is 'laid bare' to the thinker."[19] This mental contact can also be captured by the technical concept of *acquaintance*[20]: the intuition of obviousness is an intuition of acquaintance.

[19] Goff (2015, 122).
[20] See Russell (1911; 1912) for the concept, and Duncan (2021) for a contemporary presentation.

The careful reader noticed one thing. The intuition of obviousness is a *psychological* appearance: having the intuition means having a disposition to believe certain things. Interestingly, the content of this intuition is itself spelled out *in terms of appearances*. Phenomenal consciousness psychologically *appears* introspectively obvious when we are disposed to believe something *about its way of appearing*, namely, that its introspective appearance and its reality coincide. We can then ask: Which concept of appearance does this last occurrence of "appearance" express? In what sense of "appearance" do we tend to believe that phenomenal experiences' introspective *appearance* and reality coincide? This question is crucial. I answer it much later, in Chapter 6.

For now, why think that we have an intuition of obviousness? First, because many people *believe* that phenomenal consciousness is introspectively obvious. Consider recent examples in analytic philosophy:

> To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of a pain *is* not to have a pain [. . .] in the case of mental phenomena there is no "appearance" beyond the mental phenomenon itself.[21]

> The idea of moving from appearance to reality seems to make no sense [in the case of consciousness].[22]

> But where the existence of conscious states is concerned, you can't make the distinction between appearance and reality, *because the existence of the appearance is the reality in question*.[23]

> As this point is often put, in the case of phenomenal consciousness there is no gap between appearance and reality, because the appearance just is the reality: how the phenomenal character seems, to the agent, is how it is.[24]

Early Modern philosophers, like Descartes and Hume, agreed:

> By the term "thought," I understand everything which we are aware of as happening within us, in so far as we have awareness of it [. . .]. It is possible

[21] Kripke (1980, 152, 154).
[22] Nagel (1974, 444).
[23] Searle (1997, 121–22).
[24] T. Horgan (2012, 406).

> for me to think I am seeing or walking, though my eyes are closed and I am not moving about; such thoughts might even be possible if I had no body at all. But if I take "seeing" or "walking" to apply to the actual sense or awareness of seeing or walking, then the conclusion is quite certain.[25]

> [S]ince all actions and sensations of the mind are known to us by consciousness, they must necessarily appear in every particular what they are, and be what they appear. Everything that enters the mind, being in *reality* as the perception, tis impossible anything should to *feeling* appear different.[26]

Similarly, 20th-century phenomenologists, like Husserl and Merleau-Ponty, concurred:

> In the psychical sphere there is, in other words, no difference between appearance and Being.[27]

> [A] being that is for itself as soon as it is because appearing, and therefore appearing to itself, is its whole being—that is the being we call mind.[28]

In many of these quotes, we find the idea that one cannot distinguish between phenomenal appearance and reality. We also find the suggestion that this is something like a conceptual truth: as Thomas Nagel puts it, distinguishing phenomenal appearance and reality "makes no sense."

Moreover, if we take the components of the obviousness intuition separately, we find many more endorsements of them.

Take the infallibility of introspective appearances. The idea is often associated with Descartes, but it is older. In the late fourth century CE, Augustine of Hippo stressed that the mind "is certain of itself." While other things are known through fallible representations, the mind knows itself through "inner, non-simulated but true presence": "nothing after all is more present to [the mind] than itself."[29] For Augustine, even a skeptic cannot refute

[25] See Descartes's *Principle of Philosophy*, first published in 1644 (1985a, 195).
[26] Hume ([1739–1740] 1896, 190). There are comparable quotes by Locke, Hamilton, Brentano, C.I. Lewis, Ayer, Ewing, Malcolm, and Shoemaker (Alston 1971, 224–25).
[27] Husserl ([1910–1911] 2002, 312).
[28] Merleau-Ponty ([1964] 1968, 30).
[29] Augustine (1990, book 10, paragraph 16).

> him who says: "I know that this appears white to me, I know that my hearing is delighted with this, I know that this has an agreeable odor, I know that this tastes sweet to me, I know that this feels cold to me."[30]

Ancient Skeptics somehow agreed. For Sextus Empiricus (second–third century CE), all knowledge of objective reality is disputable. Still, statements about our "own present experience" and the way we are "affected" (*pathos*) are "immune from inquiry (*azetetos*)," "not open to dispute."[31]

Now, take the idea that experiences are self-revealing.[32] It is accepted by various contemporary philosophers, beyond those cited above.[33] Again, Descartes is often seen as the source of the idea, but partial versions of it appear earlier. For instance, Aristotle saw perception as always accompanied by some "reflective awareness" of itself, which "is both intrinsic and immediate" and yet happens "on the side," or "*en passant*."[34] Medieval philosopher John of Ripa (14th century) thought that:

> every seeing [visio, a term Ripa uses to refer to acts of cognizing in general] in which something is shown as in an image, is first cognized as an obscure (*enigmatica*) seeing of itself.[35]

Comparable views, on which conscious mental states are such that just being in them is enough to be aware of them, were arguably held by other medieval philosophers, such as William of Auvergne (13th century), Thomas Aquinas (13th century), Peter John Olivi (13th century), or Walter Chatton (14th century).[36] That experiences immediately self-present does not imply that they self-reveal: immediate self-presentations can leave out essences, or mislead. Still, immediate self-presentations seem epistemically more secure than standard perceptual presentations, and at least some medieval thinkers thought that this self-presentation of experiences was immune from certain errors, and grounded uniquely certain forms of knowledge.[37]

[30] Augustine (1993, book 3, §26).
[31] Burnyeat (1982, 26).
[32] See M. Liu (2024) for the relation between revelation and the absence of appearance/reality distinction.
[33] Goff (2017); Nida-Rümelin (2007); Strawson (2003).
[34] Caston (2002, 786–87).
[35] Lavender (2022, 204).
[36] Cory (2014; 2021). Thanks to Jordan Lavender and Therese Cory for the references.
[37] E.g., Thomas Aquinas (Cory 2014, 175, 191).

A similar idea appeared outside the West. The view that "all awareness (consciousness, experience) somehow or other involves awareness of that awareness" seems common in classical Indian philosophy, whether we look at Buddhist philosophers (Dignāga, Dharmakīrti, Śāntarakṣita) or Advaitins (Śaṅkara, Citsukha).[38] A similar view appears in Medieval Islamic Thought. For instance, the Persian Shihab al-Din al-Suhrawardi (12th century), among the most influential Islamic philosophers, thought that a unique, immediate epistemic relation holds between subjects and their experiences—the experience of pain being paradigmatic in this respect, as the mere presence of such an experience makes it perceived and known immediately by the subject who has it.[39]

To summarize, many thinkers, throughout history, thought that phenomenal consciousness *is* introspectively obvious. This suggests that it at least *appears* to be such.

1.4. The intuition of irreducibility

Our second mystery intuition is an irreducibility intuition. Phenomenal consciousness seems *irreducible*—epistemically and metaphysically.

First, *epistemic* irreducibility. We struggle to explain phenomenal consciousness in scientific terms. Cognitive science views mental phenomena as a set of complex processes in our brains, bodies, and environment, including chemical reactions, electrical activities, patterns of neuron firing, computational processes, and efficient coupling with the environment. These processes and their properties explain much about behavior and cognitive capacities, but not *why* we have phenomenal experiences, nor *these* experiences in particular. An "explanatory gap" seems to remain.[40] Trying to bridge the gap, we face the "hard problem of consciousness."[41]

On the other hand, explaining the fulfillment of cognitive *functions* associated with consciousness (e.g., explaining access-consciousness or self-consciousness) only raises so-called easy problems. These easy problems are not effortless to solve. What makes them "easy" is that our current methods, in principle, seem up to the task. For instance, a state is access-conscious when it plays a certain functional role in a cognitive system, and we think that, in

[38] Strawson (2015, n. 17).
[39] Marcotte (2006, 549–50; 2012).
[40] Levine (1983; 2001).
[41] Chalmers (1995).

principle, our current methods can explain this sort of functional roles. In contrast, we do not see how they could explain phenomenal experiences.

Two aspects of phenomenal experiences are sometimes distinguished.[42] First, they seem to have a *qualitative* aspect, making them *like* something: an olfactory experience of coffee qualitatively differs from a visual experience of red, which also qualitatively differs from an experience of green. This qualitative aspect seems *ineffable*, impossible to capture entirely in verbal terms and to communicate to someone who does not already know about it: you cannot fully explain what it is like to see red to a blind person. Second, phenomenal experiences seem to have a *subjective* aspect—to be like something *for the subject who has them*. My smelling coffee and my seeing red do not float free and subjectless: they are, in some deep sense, *for me*. Arguably, both the qualitative and the subjective aspects of phenomenal consciousness make it hard to explain, as they do not seem to be a matter of fulfilling a functional role.

Second, metaphysical irreducibility. To say more about what phenomenal consciousness's apparent metaphysical irreducibility consists in, we can first introduce *physicalism* (or *materialism*). It is the metaphysical view that everything is physical: everything that exists is nothing over and above a set of physical entities with physical properties.

Defining "physical" is tricky.[43] It can refer to concretely existent objects, states, events, and properties that can be fully described in terms of structures, the relevant structures being logico-mathematical, spatiotemporal, and causal/functional/nomic.[44] This is the definition I primarily use. In this strong 'structural' sense, physical concepts are structural concepts.

This requires a slight revision of what the conceptual independence of physical and phenomenal concepts implies. Given this independence, it remains true that, from propositions about physical properties, we cannot a priori deduce non-trivial propositions about phenomenal properties. Still, from some propositions about phenomenal properties, we *can* a priori deduce some non-trivial propositions about structural (i.e., physical) properties. For instance, from the proposition that I have a complex visual experience with four distinct colored parts instantiating some geometrical structure, I can deduce that this geometrical structure is instantiated. Still, I assume that, from propositions about phenomenal properties, non-trivial

[42] Levine (2001, 6–9); Kriegel (2005).
[43] Hempel (1980); Stoljar (2024).
[44] Chalmers (2002a); Alter (2016).

propositions about causal/functional/nomic structures cannot be deduced a priori. Causally inert phenomenal experiences are conceivable.

"Physical" does not have to be defined in this way. In a weak and negative definition, it refers to something that either *does not have* certain characteristics, such as *being mental*, or is entirely constituted by other physical things.[45]

Arguably, physical entities in the structural sense are physical in the weak sense, but not the other way around. Consider a property that is *intrinsic* (an entity has it in itself), *categorical* (it concerns what the entity is, and not what it does—i.e., it is not a causal or dispositional property), and *fundamental* (it is not constituted by, nor analyzable into, other properties). Properties like this are *fundamental qualities* or *quiddities*. Now, suppose a *non-mental* quiddity. Arguably, it is not captured by structural descriptions. It is weakly, but not strongly, physical.

In both senses of "physical," physicalism is attractive to some degree—more so in the weak than in the strong sense. What makes physicalism attractive is the predictive, explanatory, and pragmatic success of sciences that only posit physical entities. If "physical" is read in the strong structural sense, physicalism is close to the view called ontic structural realism in the metaphysics of science, on which there are only structures.[46]

Phenomenal consciousness seems to be non-physical. We have an intuition of *metaphysical irreducibility* of consciousness to the physical, sometimes also called an "intuition of distinctness."[47]

Famous anti-physicalist arguments elicit this intuition. Frank Jackson's Knowledge Argument features Mary, a neuroscientist who spent all her life in a black-and-white room. There, she learned everything there is to know about the physical processes underlying *color perception* without ever seeing colors.[48] Leaving the room, Mary sees her first red object. Does she learn something new? Jackson thought so: she learns *what it is like* to see red. Since she already knew everything about the physical processes involved in the perception of colors, the property she just learned about must be non-physical. Physicalism is false.

Other influential anti-physicalist arguments include conceivability arguments,[49] such as the Zombie Argument.[50] In David Chalmers's version,

[45] Spurrett and Papineau (1999); Montero and Papineau (2005).
[46] Ladyman (1998); Esfeld (2009); French and Ladyman (2010).
[47] Papineau (2002; 2008).
[48] Jackson (1982; 1986). For the history of similar arguments, see Nida-Rümelin and O'Conaill (2021, sec. 1). Jackson later changed his mind and embraced physicalism (Jackson 1998b).
[49] See Descartes's *Meditations*, first published in 1641 (1985b); see also Kripke (1980).
[50] Kirk (1974); Chalmers (1996).

it goes like this. We can conceive of zombies: physical duplicates of phenomenally conscious humans devoid of phenomenal experiences. However, if they are conceivable, they are metaphysically possible: they *could* exist, in a very broad understanding of "could" (roughly, God could have made them). But, if they are metaphysically possible, physicalism is false: physicalism implies that phenomenal consciousness, here assumed to exist, is physical, which implies that a perfect physical duplicate of a phenomenally conscious being is a phenomenal duplicate—making zombies impossible. Therefore, physicalism is false.

Intuitions of epistemic and metaphysical irreducibility are deeply related. As Joseph Levine puts it:

> whether we think of [the gap] as an explanatory gap or a distinctness gap, the problem is really the same.[51]

For instance, science seems able to explain physical phenomena. If the phenomenal seemed metaphysically reducible to the physical, it would also seem epistemically reducible. Conversely, it is plausible that explanations require the explanandum to be deducible from its explanans. Since, arguably, science only posits physical entities, scientifically explaining the phenomenal would mean deducing phenomenal truths from physical truths. If we could do this, zombies would become inconceivable, and we would see how Mary could know everything about the phenomenal before leaving her room. The phenomenal would cease to seem metaphysically irreducible.

So far, I have only described an intuition of irreducibility *to the physical*. However, our intuition of irreducibility is more general than that. Phenomenal consciousness seems metaphysically irreducible to anything that is substantively conceived in a way that is not straightforwardly phenomenal.

Consider the non-mental quiddity mentioned earlier. It does not amount to a structural property, and it is not physical in the strong sense. Still, we seem as reluctant to reduce the phenomenal to it as to structural properties. Worse, consider a *mental* quiddity that is *not conceived of as phenomenal*: e.g., a fundamental, intrinsic, and categorical property of having free will or being a rational subject. This quiddity is not physical in any sense. Yet, we intuitively resist reducing the phenomenal to it. So, our intuition is

[51] Levine (2007, 148).

that the phenomenal is metaphysically irreducible to anything substantively conceived in a way that is not straightforwardly phenomenal.[52] The irreducibility intuition is mostly discussed vis-à-vis the physical only because reduction to the physical is otherwise judged attractive.

The intuition of irreducibility—epistemic or metaphysical—is widely shared. Many philosophers endorse it. A substantive minority of contemporary philosophers thinks that the phenomenal is non-physical, and a majority thinks that it creates a hard problem—i.e., is epistemically irreducible.[53] Many would share the view expressed in the following striking quotes[54]:

> You speak [. . .] of this "emergence" of consciousness from matter [. . .] but surely such a feat is quite inconceivable. [. . .] You might as well assert, without further explanation, that numbers emerge from biscuits, or ethics from rhubarb.[55]

> Nobody has the slightest idea what consciousness is, or what it's for, or how it does what it's for (to say nothing of what it's made of).[56]

Many non-philosophers also think that consciousness is irreducible—at least epistemically. Most consciousness researchers (in the main part cognitive scientists) believe in the explanatory gap.[57] The belief is also common outside of cognitive science. Erwin Schrödinger (Nobel Prize recipient and one of the fathers of quantum physics), claimed that:

> The sensation of colour cannot be accounted for by the physicist's objective picture of light-waves. Could the physiologist account for it, if he had fuller knowledge than he has of the processes in the retina and the nervous processes set up by them in the optical nerve bundles and in the brain? I do

[52] Still, we have no intuition of metaphysical irreducibility of the phenomenal to things that are not substantively conceived: we easily identify a phenomenal experience with the referent of "this entity."

[53] The last PhilPapers survey polled 7685 professional philosophers; 62% of them accepted (or leaned toward) the view that there is hard problem of consciousness (Chalmers and Bourget 2020), and 32% of them accepted (or leaned toward) non-physicalism about the mind—which, we can suppose, often amounts to non-physicalism about phenomenal consciousness.

[54] Citations are borrowed from Humphrey (2019).

[55] McGinn (1993, 160).

[56] Fodor (2004, 31).

[57] A total of 166 consciousness researchers, coming from cognitive neuroscience, psychology, and philosophy, were surveyed during meetings of the *Association for the Scientific Study of Consciousness* (Francken et al. 2022).

> not think so. We could at best attain to an objective knowledge of what nerve fibres are excited and in what proportion, perhaps even to know exactly the processes they produce in certain brain-cells [. . .] But even such intimate knowledge would not tell us anything about the sensation of colour.[58]

Physicist Edward Witten (Medal Fields recipient, famous for his work on string theory, and informally voted "smartest living physicist" in 1990),[59] stated:

> I think consciousness will remain a mystery [. . .]. I tend to think that the workings of the conscious brain will be elucidated to a large extent. Biologists and perhaps physicists will understand much better how the brain works. But why something that we call consciousness goes with those workings, I think that will remain mysterious. I have a much easier time imagining how we understand the Big Bang than I have imagining how we can understand consciousness.[60]

Popular wisdom takes physics to be the most "sciencey" of all fields. If top physicists think that consciousness is epistemically irreducible, the idea is not a quirk of science-ignoring philosophers.

Looking at these recent examples, one might think that the intuition of irreducibility is a 20th-century phenomenon. It is not. The first conceivability argument for the metaphysical irreducibility of something like phenomenal consciousness arguably appeared in the sixth of Descartes's *Meditations*, first published in 1641:

> The fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God [. . .]. Thus, simply by knowing that I exist and seeing at the same time that absolutely nothing else belongs to my nature or essence except that I am a thinking thing, I can infer correctly that my essence consists solely in the fact that I am a thinking thing [. . .]. And accordingly it is certain that I am really distinct from my body, and can exist without it.[61]

[58] Schrödinger (1958, 90).
[59] J. Horgan (2014).
[60] J. Horgan (2016).
[61] Descartes (1985b, 54).

Half a century later, in 1689, Locke presented the idea at the basis of the hard problem, stating that:

> Matter, *incogitative Matter* and Motion, whatever changes it might produce of Figure and Bulk, *could never produce Thought*.[62]

In 1702, Pierre Bayle fleshed out the idea. Discussing the materialist view that the human mind is nothing but "a machinal construction and disposition of several parts of matter," and notably addressing the ideas of an anonymous philosopher (in fact, the Irish materialist John Toland), he claimed:

> I maintain that he supposes what has hitherto been inconceivable to all Mankind, if he supposes that the sole disposition of the organs of a human body should cause a Substance to become thinking, that never thought before. All that the disposition of these organs can do, is confined to local Motion, variously modified, as in a Clock (*horloge*); all the difference being only from more to less [. . .]. All you can do to [Extension] by pulling, striking, or pushing it all manner of ways, is a change of situation, the whole nature and essence of which you can fully conceive, without having recourse to any sensation (*sentiment*), and even when you deny there is any sensation (*sentiment*) in it.[63]

Leibniz, maybe influenced by his reading of (and correspondence with) Bayle,[64] wrote, a decade later, in 1714, the famous *Monadology*'s mill argument, often seen as the first version of the explanatory gap argument:

> [P]erception, and what depends upon it, is inexplicable in terms of mechanical reasons, that is through shapes, size, and motions. If we imagine a

[62] Locke ([1689] 1975, 623). Locke ([1689] 1975, 541) conceded that God could still "superadd" thought to matter.

[63] Bayle ([1702] 1735, 661, entry "Dicaearchus"). The French "sentiment" might more adequately be translated by "feeling."

[64] Leibniz had avidly read the book by Bayle that features these lines—the second edition of Bayle's dictionary, in French, printed in December 1701 (Leibniz 1887, 63)—in the summer of 1702. In 1702, Leibniz was in discussion with Toland and Queen Sophie Charlotte (in person) as well as Bayle (in writing) on the topic (Dagron 2009). In a letter to Bayle of the Fall 1702—a letter he drafted four times but never sent (Leduc et al. 2015, 188–95)—Leibniz agreed with Bayle to reject Toland's materialism for the same reasons as Bayle's, adding a *mill* analogy to Bayle's clock (Leibniz 1887, 68), and writing in ways that prefigure the *Monadology*'s argument. This is the earliest use of the mill analogy by Leibniz for this purpose—or the second earliest, since a roughly contemporary letter to Sophie Charlotte is hard to date exactly (Leibniz 2011, 258–59), but even this second letter was likely written after Leibniz has read Bayle's 1702 dictionary. An influence of Bayle on Leibniz seems plausible.

> machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters a mill. Assuming that, when inspecting its interior, we will find only parts that push one another, and we will never find anything to explain a perception.[65]

In the 19th century, we start seeing formulations of the hard problem very similar to ours. Biologist Thomas Huxley is often cited:

> How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nerve tissue, is just as unaccountable as the appearance of the Djinn when Aladdin rubbed his lamp.[66]

The claims of the Irish physicist John Tyndall, a few years later, are also often mentioned:

> The passage from the physics of the brain to the corresponding facts of consciousness is unthinkable.[67]

Some view Huxley, Tyndall, as well as the slightly posterior Emil du Bois Reymond,[68] as the modern originators of the hard problem, after a century of (relative) post-Leibnizian disinterest.[69] They are also seen as the first proponents of *mysterianism*, the view that consciousness, though not objectively unexplainable, will always remain so for us.[70] However, these ideas were not highly original at the time. A few years before Huxley, Wilhelm Griesinger, the father of modern German psychiatry, stated both the hard problem and the mysterian option:

> How a material, physical process going on in the nerve fibers or the ganglion cells could become a representation, an act of consciousness, is completely incomprehensible (*unbegreiflich*) [. . .]. If we knew everything that happens in the brain during its activity, if we could see through all

[65] Leibniz (1989a, 215).
[66] Huxley (1866, 193).
[67] Tyndall (1871, 119). The claim was made in an 1868 address.
[68] Bois-Reymond (1872).
[69] "Leibniz did not elaborate and there was not a great deal of further discussion around that time. For really definitive statements of the hard problem of consciousness, we have to wait until the 1860s [. . .]. Huxley was first" (Chalmers 2020c, 232).
[70] Tennant (2007).

> the chemical, electrical, etc. processes, down to their last detail—what use would it be? All the oscillations and vibrations, everything that is electrical and mechanical, are still not a state of the mind, a representation. How could they become such—this riddle (*Räthsel*) will probably remain unsolved until the end of time; and I believe that, if today an angel came down from heaven and explained it all to us, our intelligence would not even be able to comprehend (*begreifen*) it.[71]

Griesinger made these claims in the introduction of the second edition of his psychiatry manual, as a bit of elementary propaedeutic, not as the result of deeply original thought. Others made similar points between Leibniz and him. Séverin de Cardaillac, a then-influential French philosopher, pointed out the explanatory gap in 1830, stressing

> the vanity of the attempts that have been made, and are still being made, to discover, in man's physical nature, in this admirable apparatus of nerves, muscles, sensory and motor organs, the explanation of the phenomena of sentience, intelligence and activity of man.[72]

Cardaillac insisted that it is "impossible," even after the most thorough study of (cerebral) "impressions" on the one hand, and (conscious) "sensations" on the other hand, to

> bridge the depth of the gap (*abîme*) that exists between them.[73]

He rammed it home:

> The physiologist can study and know, through exact anatomy, all the details of the system of sensory organs. He will be able to guess the means of communication between the external impression and the brain. It might even be that, with more perfect instruments, he will discover the particular impression, inside the brain, that is necessary for each

[71] Griesinger (1861, 6, my translation).
[72] De Cardaillac (1830, 36, my translation).
[73] The French "abîme" means "gap," "abyss," or "chasm." This is, to the best of my knowledge, the oldest use of a term like "gap" to qualify the relation between consciousness and the brain. I have not seen it cited anywhere in the recent literature: John Tyndall and Emil du Bois-Reymond are usually seen as the earliest users of this terminology.

> sensation; but what will he find there? Some movement, some configuration. As for the sensation itself, it eludes the scalpel. Between the last impression and the sensation, there is infinity. The last impression will be the end goal of all the physiologist's discoveries [. . .]. Certainly, the physical apparatus is linked, in man, to sentience, by a tight chain [. . .] but they are of such a different nature that the link between them will forever escape our research.[74]

Cardaillac, like Griesinger, presented these claims as introductory philosophical platitudes. The idea that phenomenal consciousness is epistemically irreducible has been commonplace in the West, at least from Leibniz onwards. When it was not discussed explicitly during this time, it is probably because it seemed too banal to deserve a defense.

Many philosophers and scientists, past and present, believed that phenomenal consciousness *is* irreducible—epistemically and/or metaphysically. This strongly suggests that it at least *seems* irreducible. We have an intuition of irreducibility.

1.5. The intuition of undetectability

Phenomenal experiences seem introspectively obvious, but outside of introspection, they seem impossible to detect directly. We have an *intuition of undetectability*.

How can you tell that your best friend is phenomenally conscious and not a zombie? You cannot take a peek directly into her inner life. She behaves and speaks as if she has experiences, but a zombie would do the same. You *feel* that she has experiences: seeing her falling off a cliff and screaming while holding her foot, you empathize with her and feel her pain. Still, a zombie would elicit the same empathy.

So, we can directly observe the behavior of others and their superficial physical properties, but not their phenomenal experiences. When it comes to humans, the difficulty seems intractable in principle, but innocuous in practice. We just assume that, as humans are very similar behaviorally and physically, they are also similar phenomenally. However, consider other creatures.

[74] De Cardaillac (1830, 37–38, my translation).

Which animals are phenomenally conscious? How can we know? They cannot tell us. Many are behaviorally, functionally, and physically similar to us in some ways—but are they similar enough and in the right way? We intuitively grant phenomenal consciousness to apes and dogs, but what about mice? Squids? Grasshoppers? Ants? Worms? Starfish? Oysters? Beyond animals, what about plants? Lab-grown models of human brain parts ("cerebral organoids")? What about vegetative patients? Early-stage fetuses? Robots? Artificial neural networks? What about tightly organized groups—beehives, modern states, or companies? All these are behaviorally, functionally, and physically similar to adult humans in *some* respects. As phenomenal consciousness seems impossible to detect directly, establishing which beings are phenomenally conscious is challenging.

It also seems hard to know *which* phenomenal experiences others have. Suppose my friend is not a zombie. Still, how can I know if, looking at a *Rosso Corsa* Ferrari, she has, like me, an experience of red, rather than a different experience? Could her color experiences be inverted compared to mine?

Again, we often assume that behavioral and physical similarity indicate phenomenal similarity, but this does not always solve the issue. People in similar situations sometimes behave similarly, but describe their experiences differently. Some claim to have *sui generis* phenomenal feelings associated with abstract cognitive processes—thinking an imageless thought, understanding a sentence—while others deny it.[75] Some report rich and detailed experiences during brief visual presentations, "overflowing" what they can access. Others do not.[76] Are these groups really having different experiences? Do they just use words differently? Are some right and others wrong? As we cannot directly detect their phenomenal experiences, the issue is hard to settle.[77] Again, things seem worse when focusing on creatures who cannot describe their experiences. Granting them phenomenal consciousness, what is it like to be a bat, a squid, an oak, an artificial neural network? We seem unable to know.

This is the intuition of undetectability: it seems that, outside of our own case, we cannot directly detect phenomenal experiences.

[75] Bayne and Montague (2011).
[76] Cova et al. (2021).
[77] Schwitzgebel (2011).

1.6. Battleground Consciousness

Phenomenal consciousness seems obvious, irreducible, and undetectable. Daniel Dennett summed it up best: "Consciousness is both the most obvious and the most mysterious feature of our minds."[78] These mysteries of consciousness have been intensely debated. Radically conflicting views have emerged. It helps to get an idea of their diversity.

Take the apparent irreducibility of phenomenal consciousness. 'Mysterians' think the explanatory gap reveals no metaphysical gap—no dualism of any sort—but still cannot be bridged. The phenomenal will always remain unexplainable *for us*.[79] Others think it is metaphysically irreducible. They embrace dualism,[80] or neutral monism.[81] Others contend the phenomenal is explainable by yet unobserved physical effects occurring in the brain.[82] Most believe consciousness is an ordinary and explainable brain process, but they disagree on *which one*,[83] and of course, they must still explain why the phenomenal (falsely) appears irreducible.[84]

Take the obviousness of phenomenal consciousness. Some think phenomenal consciousness is indeed uniquely obvious.[85] Others defend a watered-down version of this unique obviousness, compatible with physicalism.[86] Yet others deny any special epistemic relation to phenomenal consciousness.[87]

Finally, the apparent undetectability of phenomenal consciousness leads to deep disagreement, e.g., regarding its distribution across beings. Some embrace panpsychism: they think phenomenality is present at the fundamental level.[88] Aside from panpsychism, many options have been defended, from the most liberal to the most conservative. Considering only the biological realm, some researchers reserve consciousness to primates,[89] while

[78] Dennett (1987).
[79] McGinn (1989). Griesinger, Huxley, etc., are early proponents of the view.
[80] Jackson (1982); Chalmers (1996).
[81] Maxwell (1979); Chalmers (2015); Goff (2017).
[82] Hameroff and Penrose (2014).
[83] See Seth and Bayne (2022) for a review. More in Chapter 8.
[84] This is the meta-problem of consciousness (Chalmers 2018), to which I come back in the next chapter.
[85] Chalmers (2003); Gertler (2012); Goff (2015); Giustina (2022).
[86] Balog (2012a); Coleman (2019).
[87] This is the case of illusionists, but some non-illusionists radically deny this too (Schwitzgebel 2011).
[88] Chalmers (2015); Roelofs (2019).
[89] Bermond (2001).

others suggest that all animals,[90] or even all living beings,[91] are phenomenally conscious.

Consciousness studies is a Kantian *Kampfplatz*: a battleground, with little to no agreement on most significant issues. The mysteries of consciousness run deep.

1.7. Mystery intuitions are shared and compelling

This leads me to the most significant presuppositions of this book. *We should take the mysteries of consciousness seriously*.

I do not mean to presuppose that we have a genuinely mysterious form of consciousness, whatever this means. I only presuppose that phenomenal consciousness *really seems* mysterious: obvious, irreducible, undetectable. It does so in a way that is not merely idiosyncratic, weak, and fragile: our mystery intuitions are *widely shared*, *robust*, and *strong*. (To say, in one word, that they are both robust and strong, I say that they are *compelling*). Moreover, these intuitions are *conceptual*: they stem from the way our phenomenal concepts characterize phenomenal experiences.

I first argue that our mystery intuitions are shared and compelling. Later, I argue that they are conceptual. Objections to these presuppositions are examined later, in Chapter 7.

A. Mystery intuitions are widely shared

Our mystery intuitions are dispositions to believe. These are specifiable along various parameters, such as *distribution*, *robustness*, *strength*, and *origin*. Focus first on *distribution*.

Dispositions to believe are differently distributed among subjects. Almost everyone is disposed to believe they exist, but few are disposed to think that their name is François. I suppose that *mystery intuitions* are widely shared. The vast majority of adult humans have phenomenal concepts and, *when* they have other required background concepts and engage in sufficient reflection, they *tend* to believe that phenomenal experiences are mysterious: obvious, irreducible, undetectable.

[90] Andrews (2024).
[91] Thompson (2022).

I mention other concepts to account for the fact that having an actual tendency to believe that phenomenal consciousness is irreducible to the physical, or unexplainable, requires the right concepts of explanation, or physical. I mention sufficient reflection to account for the fact that many people never think about these issues or do it unreflectively, and thus never contemplate these problems. Finally, even someone who actually tends to believe that the phenomenal is mysterious can resist the tendency and avoid the belief.

Given these precautions, we could formulate otherwise the view that mystery intuitions are widely shared. We could use "intuition" to refer to *active tendencies to believe*, and claim that mystery intuitions in this sense are *not* shared, even if *key sources* of these intuitions are.[92] The alternative between this formulation and mine is verbal. For reasons of simplicity, I stick to the formulation on which *intuitions* themselves underlie active tendencies to believe and are widely shared, so that many individuals might have the intuitions but lack corresponding active tendencies.

Why think that mystery intuitions are widely shared? On a first approach, because, as I just showed, many thinkers from various cultural, linguistic, and disciplinary backgrounds thought phenomenal consciousness *is* mysterious. This suggests an underlying shared intuition. Against this, an objection is that the examples I gave earlier are not diverse enough. Some mystery intuitions seem to be reported only in a few cultures and historical periods. This objection is strongest for the intuition of irreducibility, and I focus on it now.

Reports of irreducibility intuitions mentioned earlier came from modern Western thinkers. Such reports are rare before the 17th century.

For instance, the hard problem of consciousness does not seem to be ancient.[93] We find few (if any) convincing statements of something like the hard problem in the Ancient world. Of course, most prominent Ancient philosophers believed in the immateriality of (parts of) the human mind (think of Plato's *Phaedo* or Aristotle's *De Anima*), but their arguments did not focus on the mind's phenomenal features.

Some of the most convincing Ancient anticipations of the basic idea powering the hard problem are to be found in criticisms of atomistic theories. For instance, here is Cicero, in the first century BC[94]:

[92] Chalmers (2020c, 228) thus distinguishes *intuition universality* and *source universality*.
[93] Matson (1966).
[94] The tempting view that Jesus Christ himself stated the hard problem in the (apocryphal) Gospel of Thomas must probably also be rejected: "Jesus says: 'If the flesh came into being because of the spirit, it is a wonder. But if the spirit (came into being) because of the body, it is a wonder of wonders'.

> Even if [atoms] existed, which they do not, they might perhaps be able to strike one another and be moved by mutual collisions, but they would not be able to regulate, to imagine, to give color, to animate.[95]

See also Galen, in the second century CE:

> A single body capable of sensation cannot be produced from many which are incapable of sensation.[96]

Moving forward in time, the hard problem of consciousness does not seem medieval either.[97] Looking outside the West yields similarly disappointing results: at best, we find only hints of something like the hard problem, for instance, in Dharmakīrti's argument (seventh century CE) against the physicalist school of the Cārvāka, here interpreted by his follower Prajñākaragupta:

> If the effect, [i.e.] consciousness, arises from these alone, [i.e.] from the great elements, without coming from another life, why does not everything arise having the form of a living being? [. . .] Why does the special transformation not [arise everywhere]?[98]

Does all of this show that irreducibility intuitions are not widely shared? I grant that irreducibility intuition reports are scarce outside the modern West, but two factors naturally explain it, even assuming that these intuitions themselves are widely shared.

First, as I noted, irreducibility intuitions only generate active tendencies to believe (and therefore, reports) when other concepts are mastered. The intuition of metaphysical irreducibility requires the right concept of the

Yet I marvel at how this great wealth has taken up residence in this poverty" (Patterson et al. 2011, 8). Thanks to Nick Humphrey for the reference.

[95] Cited by Glidden (1979, 305).

[96] Cited by Chalmers (2020c, 230). Galen's ideas might have had an influence on modern formulations of the problem of consciousness. Bayle, whose writing and correspondence with Leibniz arguably played a role in the first elaborations of the explanatory gap, knew of Galen's argument, and claimed that atomistic explanations of the mind fell prey to Galen's "thundering objection," of which he cited a Latin paraphrase by the 17th-century atomist Pierre Gassendi (Bayle [1702] 1735, 779, entry "Epicurus").

[97] King (2007b); see Lavender (2024) for a dissenting opinion.

[98] Franco (1997, 106). Citation taken from Chalmers (2020c, 231).

physical, making the physical at least fundamentally non-qualitative and not mental, and maybe also exhausted by the structural. However, this concept of the physical is not widely shared. It only became dominant in the West in the 17th century. Earlier, the West and the Muslim world mostly used an Aristotelian concept. For Aristotle, the physical world includes fundamental sensible qualities such as whiteness and warmth,[99] and is animated by souls. In the 17th century, a concept of the physical as something exhausted by mathematical descriptions emerged in the work of Galileo and Descartes. This elegantly predicts that active tendencies to see consciousness as metaphysically irreducible should emerge in the West during the 17th century, even if the underlying intuition, in my sense, is more widely spread.[100]

Second, an intuition only leads to an active tendency to believe given sufficient reflection, and an active tendency to believe only leads to reports when the context calls for it. So, we should expect the irreducibility of consciousness to be frequently discussed only when the context makes it worth reflecting on it, and talking about it. This will typically happen when *materialism* about the mind is taken seriously.

This naturally explains the pattern of reports described earlier.[101] Cicero and Galen were answering atomists. Dharmakīrti was answering the Cārvāka school. Bayle and Leibniz felt the need to answer John Toland—as well as atomists like Gassendi, or other materialists like Hobbes. Reports became more numerous and detailed as materialism became more influential. Cardaillac, who insisted that consciousness was unexplainable in early 19th-century France, was answering the influential *Idéologues* like Cabanis who claimed that the mental "is nothing but the physical, considered under certain particular point of view,"[102] or Broussais, the most influential French physician at the time, and a radical materialist.[103] Griesinger was probably reacting to the materialism of Büchner, Vogt, or Moleschott, which gained traction in German academia in the 1850s and led to the "controversy of materialism" (*Materialismus-Streit*).[104] This growing influence of materialism

[99] Ganson (1997).
[100] A standard candidate explanation of why the mind–body problem is not medieval (King 2007b, 204).
[101] See also Chalmers (2020c, 233).
[102] Cabanis (1805, 40, my translation).
[103] Broussais (1828).
[104] Bayertz et al. (2007).

culminated in the "materialistic '60s,"[105] motivating Huxley's, Tyndall's, and du Bois-Reymond's statements. So, even if we assume that irreducibility intuitions are widely shared, we should expect irreducibility claims about the phenomenal to be rare when non-materialist views are the default, and to become more common as materialist views gain popularity. This seems to correspond to the observed patterns of reports.

B. Mystery intuitions are compelling

Dispositions to believe also differ in terms of *robustness*: they can be destroyed more or less easily. My disposition to believe that my new colleague is arrogant is *fragile*. It disappears when I understand her apparent arrogance is disguised timidity. My dispositions to believe that 2 + 2 = 4, or that simultaneity is absolute, are very robust. I can be convinced that simultaneity is not absolute, but I still tend to believe it is.

One way of being robust is to be *cognitively impenetrable*, and notably resistant to the acquisition of contradictory beliefs. My perceptual disposition to believe that there is a table in front of me disappears when I move my head or close my eyes, but as long as I look at the table, it remains. I might be convinced I face a trompe-l'oeil or suffer from hallucination, but the tendency does not disappear. Perceptual dispositions to believe are typically robust, at least in the sense that they are cognitively impenetrable.

I suppose that mystery intuitions about phenomenal consciousness, conditional on the use of phenomenal concepts, are robust, and notably cognitively impenetrable. Why believe this? Because many do *not* endorse these intuitions but still report them.[106]

This is striking for the intuition of irreducibility. Consider what two full-blown materialists, Daniel Dennett and David Papineau, wrote:

> I know the intuition [that consciousness does not reduce to something physical] well. I can feel it myself. [. . .] I feel it, but I don't credit it.[107]

[105] Russell (1950).
[106] *Pace* Rosenthal (2019a; 2022), for whom people usually only report *intuitions* closely correlated with the *opinions* they (or the person eliciting the intuitions) hold.
[107] Dennett (2018, 456).

> This [anti-materialist] intuition continues to operate even in those, like myself, who are otherwise persuaded that there are good arguments for materialism, and stops us really believing the materialist conclusion.[108]

The same applies to the intuition of obviousness. Many who did not endorse it still reported its strength. David Lewis denied that experiences are self-revealing—something he thought incompatible with his materialism—but admitted that it "seems obvious" that they are.[109] Christopher Hill denied the absence of appearance/reality distinction about phenomenal consciousness, but granted that:

> from the perspective of folk-psychology [. . .] it is impossible to draw an appearance/reality distinction with respect to our *experiential* access to qualia.[110]

All this suggests that mystery intuitions are robust, and notably cognitively impenetrable.

Finally, dispositions to believe differ in *strength*. The robustness of a disposition concerns its conditions of existence, but its strength concerns its conditions of manifestation: a strong disposition leads to belief in a wider range of situations than a weak one. My dispositions to believe that 2 + 2 = 4, or that I exist, are extremely strong: I have these beliefs in virtually all situations. Strength and robustness dissociate. If I suffer from mild paranoid ideation, I have a robust but weak disposition to believe people hate me. A teenager going through a radical phase has a strong yet fragile disposition to form extreme beliefs.

I suppose that mystery intuitions, conditional on the use of phenomenal concepts, sufficient reflection, and the possession of the appropriate other concepts, are really quite strong. Why? Because many thinkers—but not all!—have endorsed these intuitions. Besides, even those who did not endorse them often recognized their pull.

To summarize, mystery intuitions are widely shared, but also compelling (robust and strong).

[108] Papineau (2002, 95). Similar statements abound (Balog 2012b; Hill 1997; Loar 1997; Tye 2003).
[109] Lewis (1995, 468–69).
[110] Hill (2009, 102). Here, "experiential access to qualia" means something close to what I call "introspective appearances of phenomenal experiences."

1.8. Mystery intuitions have a conceptual origin

Finally, dispositions to believe—including intuitions—can have various *origins*, or *sources*. Some intuitions have a conceptual origin. They follow from the *characterization* borne by our concepts. I say that a characterization-as-X is borne by a concept C referring to As when mastering C makes it a priori that As are X. Subjects mastering C, merely by appropriately reflecting with the concept, strongly and robustly intuit that As are X. Intuitions following from such characterizations are *conceptual*.

Some conceptual intuitions follow from *explicit* characterizations borne by concepts—from the accepted explicit definitions of these concepts. We intuit that a blue chair is a chair, because the composed concept of a blue chair explicitly characterizes its referent as a chair: "blue chair", by definition, refers to a chair that is blue. When a conceptual characterization is explicit, two things happen. First, a competent user of the concept can normally articulate it. Second, negating this characterization (e.g., judging that a blue chair is not a chair) leads to noticeable formal contradictions. Intuitions following from the explicit characterizations of concepts are *explicit* conceptual intuitions. Arguably, they are intuitions of analytical truths. Setting aside deliberately composed concepts, it might be that very few concepts bear explicit characterizations.

Other conceptual intuitions stem from *implicit* characterizations borne by concepts. Take the concepts of colors as properties of surfaces. Plausibly, they characterize their referents so that a single surface cannot fully have two different colors at the same time. For someone who masters color concepts, it is a priori that, say, no surface can be both fully pink and fully green, or that a fully black cube cannot have yellow dots. However, the characterization is implicit. It is *not* the case that any competent user of the concepts can articulate it: it takes reflection. It is also not the case that its negation leads to noticeable formal contradictions: the idea that a fully black cube has yellow dots is not formally contradictory.

Take another example. Consider our concepts of knowledge, belief, truth, and justification. Plausibly, taken together, they characterize their referents so that some justified true beliefs (JTBs) are not knowledge. It is a priori (knowable just by reflecting on Gettier cases) that some JTBs are not knowledge. Yet, the characterization is implicit. It is not the case that any competent user of the concept of knowledge can articulate the characterization that makes it a priori that some JTBs are not knowledge. *Experts* disagree on how

to articulate it. Besides, no formal contradiction results from negating this characterization; otherwise, the JTB analysis of knowledge would be contradictory and would have been judged obviously false by everyone even before Gettier.[111]

Historically, many saw the intuitions following from the *implicit* characterizations of concepts, like the one about colored surfaces, as intuitions of synthetic a priori truths,[112] or material a priori truths.[113] I do not take a stance on the analytic/synthetic distinction, on whether intuitions following from implicit characterizations of concepts always yield truths, or on whether these truths are analytic or synthetic.

At any rate, the idea that *some* conceptual characterizations are implicit is widespread in contemporary views of concepts. For instance, in Frank Jackson's view, concepts get their identity through their place in networks constituted by folk theories, which are partly explicit, but also mostly implicit, and by reflecting on cases, we can "tease out of us" the "implicit bits."[114] In Christopher Peacocke's view, implicit conceptions are part of what constitutes the possession of certain concepts.[115] In Georges Rey's view, our deployment of concepts is governed by tacit sub-doxastic rules that are not straightforwardly accessible and remain very hard to articulate.[116] Few views of concepts preclude the existence of implicit characterization, except maybe conceptual atomism.[117] The view that there are implicit characterizations also seems anyway needed for *conceptual analysis* to be possible and non-trivial.[118]

Scenarios contradicting explicit characterizations are both negatively and positively inconceivable. I cannot *negatively* conceive of a blue chair that is not a chair: this leads to a formal contradiction. I cannot *positively* conceive of it either. Scenarios contradicting merely implicit characterizations are *negatively* but not *positively* conceivable. The idea of a surface that is both fully pink and fully green, or the view that all JTBs are knowledge, does not lead to formal contradictions, but I still cannot form a

[111] Chalmers and Jackson (2001, 320–23).
[112] Hanna (2016); Steup and Neta (2020); B. Russell (2020).
[113] Husserl ([1900–1901] 2013, 19).
[114] Jackson (1998a, 130).
[115] Peacocke (1998b; 1998a, 140).
[116] Rey (1993; 1998).
[117] For a systematic presentation of views of concepts, see Laurence and Margolis (1999); Margolis and Laurence (2023).
[118] A kind of implicit/explicit distinction is mobilized to account for the possibility of informative conceptual analyses in McGinn (2011).

positive coherent conception of such a surface, nor of Gettier's JTBs being knowledge.

Many other intuitions do *not* have a conceptual origin. Take our intuitions that contact with decaying corpses is a health hazard or that snakes are dangerous. Plausibly, they are shared and compelling. They stem from innate adaptive tendencies. Still, they are not conceptual. Our concept of decaying corpses does not characterize them as unhealthy. Our concept of snakes does not characterize them as dangerous. One can master these concepts and have the intuitions without the corresponding claims being a priori, and while conceiving, negatively *and* positively, of healthy decaying corpses and harmless snakes. Similarly, the incorrect but shared and compelling intuition that 1 kg of lead falls faster than 1 kg of feathers in the void does not follow from the characterizations borne by our concepts; nor does our incorrect but shared and compelling intuition that, if a woman is outspoken, bright, majored in philosophy, and cares about social justice, she is more likely to be a bank teller active in the feminist movement than a bank teller.[119]

Come back to mystery intuitions. I presuppose that they have a conceptual origin and stem from *implicit* characterizations borne by phenomenal concepts. Why? Because, while denying that phenomenal consciousness has the right mysterious features leads to no formal contradictions, there is still something unintelligible about it.

Consider introspective obviousness. There is no formal contradiction in the idea of a phenomenal experience whose phenomenal essence is not introspectively presented to the subject who has it. However, I do not think I can conceive *positively* of it. When I try to imagine someone who phenomenally experiences pain and yet is not introspectively presented with this feeling and what makes it a painful one, I struggle. I can imagine someone feeling pain without *judging* that they feel pain, but it is hard to imagine someone who feels pain without being introspectively presented with the painful feeling. I also struggle to imagine someone introspectively presented with a painful feeling, but who does not feel pain. Descriptions of these situations are not formally contradictory, but they barely make sense to me. Many others cited above (Nagel, Kripke, Merleau-Ponty, etc.) apparently agree. They find it impossible to distinguish between the introspective appearance of a phenomenal experience and its reality.

[119] This is the so-called conjunction fallacy (Tversky and Kahneman 1983).

Consider irreducibility. The idea that phenomenal experiences are nothing over and above certain brain processes is not formally contradictory.[120] Still, I find it hard to comprehend. I struggle to form a coherent positive conception of it. Others agreed. Bayle, cited above, found the idea that feelings are just material changes "inconceivable." C.D. Broad found the view that sensations are nothing but certain brain activities "plainly nonsensical" and "absurd."[121] Joseph Levine stressed that the view that phenomenal properties are identical to neurological properties "seems unintelligible."[122] Nagel claimed that:

> physicalism is a position we cannot understand because we do not at present have any conception of how it might be true.[123]

As for the intuition of *epistemic* irreducibility, while there is no contradiction in the idea that consciousness is physically explainable, the concept of such an explanation still seems hard to grasp positively. Consciousness appears to us to be "unaccountable" (Huxley); a "riddle" which will remain "unsolved until the end of time" (Griesinger), and whose explanation in physical terms is "unthinkable" (Tyndall).

Finally, the same applies to undetectability. I struggle to represent to myself a situation in which I can *directly*, without mediation, detect someone else's phenomenal experiences, without these experiences becoming *my* experiences. Still, this description is not formally contradictory.

So, mystery intuitions do not resemble *explicit* conceptual intuitions. Indeed, descriptions contradicting mystery intuitions are negatively conceivable and not formally contradictory. Moreover, not everyone who masters phenomenal concepts can articulate the apparent mysteries of consciousness: philosophical reflection is required.

At the same time, mystery intuitions do not resemble *non-conceptual* intuitions. Situations contradicting non-conceptual intuitions are positively conceivable on careful reflection. We can imagine healthy decaying corpses

[120] "Anti-zombies" (Frankish 2007) or "shombies" (R. Brown 2010), creatures who are phenomenally conscious and yet entirely physical, are describable without contradiction.

[121] Broad (1925, 622–24).

[122] Levine (2007, 148). See Cutter (2022) for the view that the reducibility of the phenomenal to the physical is *positively* inconceivable.

[123] Nagel (1974, 446). For Nagel, extending our understanding of the physical could render physicalism intelligible, which is compatible with the idea that phenomenal concepts (implicitly) characterize the phenomenal as non-physical *given our concept of the physical*.

and harmless snakes. We clearly understand why an outspoken woman who majored in philosophy is always more likely to be a bank teller than a feminist bank teller.[124] However, we struggle to clearly and positively represent to ourselves a phenomenal experience that is nothing but a brain process. We struggle to distinguish between introspective appearance and phenomenal reality.

Mystery intuitions most resemble implicit conceptual intuitions. Therefore, I presuppose that this is what they are. The presupposition is debatable. I discuss it in Chapter 7. For now, remember that accepting the existence of certain conceptual characterizations as mysterious is ontologically cheap. That As conceptually seem F is a fact about our concept of As. It does not entail that As exist nor that F is instantiated. We might have conceptual intuitions that dragons are scaly and breathe fire, or that ether is a propagation medium for luminous waves. Still, no scaly creature breathes fire, and light does not need a medium to propagate. Corresponding concepts are simply not satisfied.

1.9. Intuitions of significance

Finally, phenomenal consciousness seems *significant*. Intuitions of significance might not be all shared, compelling, or conceptual, but at least *some* of them are strong and reasonably shared (more on this in Chapters 9 and 10).

First, phenomenal consciousness seems highly *epistemically* significant, and notably key to epistemic justification—our possession of reasons to believe things about reality.

For instance, I believe that there is a coffee cup in front of me. I do so because I *see it* and *feel it*—because I experience it. My phenomenal experience of the cup, arguably, justifies my belief. More generally, when wondering about what justifies us to believe various things about the world, we often mention our phenomenal experiences. They seem to be our epistemic window on non-phenomenal reality.

We also seem to have a unique epistemic relation to phenomenal experiences themselves. Consider the coffee cup again. If there is anything I have a good reason to believe at the moment, it might be, precisely, that I am visually experiencing a cup of coffee in front of me. When wondering about

[124] For more, see Kammerer (2019b).

the things we are most justified in believing, we often mention salient aspects of our own present phenomenal experiences. Introspection, apparently, gives us peculiarly good justifications to form beliefs about our experiences.

On top of this intuition of epistemic significance, we have intuitions of *ethical* significance. Phenomenal consciousness appears key to *well-being* and *moral status*.

First, consider well-being. A subject has a high degree of well-being (or welfare) when things are going well for the subject—when the situation is good for the subject. In these cases, we can also say that the situation has *intrinsic prudential value* for the subject: the situation is such that, in itself, it is good for the subject. In contrast, a situation has mere *instrumental* prudential value when it is merely able to bring about intrinsic prudential value. Well-being comes in degrees; it can be positive or negative.

Phenomenal consciousness seems key to well-being.[125] This is notably intuitive with *valenced* phenomenal experiences, which are attractive (positive valence) or aversive (negative valence). Pleasure, joy, bliss, and excitement are examples of positively valenced experiences. Pain, sorrow, and distress are examples of negatively valenced experiences. Valenced phenomenal experiences seem crucial to well-being: they appear to be "what makes life worth living."[126] Without these experiences, it is unclear whether our lives could still be good or bad for us.

Second, consider *moral status*. A being has a certain moral status (or moral standing) when, intrinsically, it deserves a certain kind of moral consideration: when, because of what it is in itself, it has certain moral rights, or dignity, etc.

Sentience in the narrow sense, the capacity to have valenced phenomenal experiences, seems key to moral status. For instance, it is plausible that the property of having a certain degree of well-being, i.e., of being a subject of welfare—*welfare subjectivity*—is necessary and sufficient for moral status. It is also plausible that the degree of well-being a creature can attain influences its degree of moral status. If so, since valenced phenomenal experiences are intuitively key to well-being, sentience is intuitively key to moral status. However, the importance of sentience for moral status is also directly intuitive. We think that humans, apes, dogs, etc., matter morally much more than trees, rocks, cars, and laptops, notably because we think that the first, but

[125] See Kriegel (2019) for other varieties of value and their links to consciousness.
[126] To paraphrase Wilfrid Sellars (Dennett 1991, 383).

not the second, are sentient. We urgently wonder whether fetuses, vegetative patients, fish, brain organoids, robots, and neural networks are sentient, to determine how we should treat them—because, intuitively, we think that sentience is key to moral status.

* * *

In this chapter, I defined "phenomenal consciousness." I showed that it seems mysterious—obvious, irreducible, and undetectable—and highly significant. Now, I turn to illusionism.

2

Illusionism About Phenomenal Consciousness

Things are not always what they seem; first appearances deceive many.

—Phaedrus

2.1. Illusionism about phenomenal consciousness

My view is a version of *illusionism* about phenomenal consciousness. Illusionism conjoins two claims. First, the *eliminative* claim: phenomenal consciousness does not exist. It does not exist more than sirens, dragons, witches, mana, phlogiston, the divine rights of European kings, the minds of angels, and the powers of djinns. Second, the *appearance* claim: it seems to exist. We are under the illusion of phenomenal consciousness.

Illusionists accept both claims. Accepting the eliminative claim makes you an *eliminativist*; rejecting it makes you a *phenomenal realist*. Here, "eliminativism" means *entity* eliminativism, not *discourse* eliminativism: it is the view that phenomenal consciousness does not exist, not that we should stop talking about it.[1] All illusionists are eliminativists. Since almost everyone accepts a version of the appearance claim, most eliminativists are illusionists, and most opponents to illusionism are realists.

The appearance claim says that phenomenal consciousness seems to exist. "Seems," here, is to be read psychologically. The claim is that we are disposed to believe that phenomenal consciousness exists. More precisely, in my view, we are *introspectively* disposed to believe that we have phenomenal experiences. These dispositions are widely shared and compelling.

[1] Irvine and Sprevak (2020).

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0003

The appearance claim cannot be about *phenomenal* appearances. If it were, illusionism would be contradictory: without phenomenal experiences, nothing phenomenally appears. It can be about *epistemic appearances*: while this is not my primary reading, I think that in this reading too, the claim is correct: phenomenal consciousness *epistemically* seems to exist, in the sense that we have some prima facie reason to believe it exists—even if it does not. In Chapter 10, I clarify illusionism's implications for epistemic justification, as well as the relation between the epistemic and psychological appearance of phenomenal consciousness.

2.2. Varieties of illusionism

A. Some history

Illusionism is a radical view. Contrary to other views of consciousness, it lacks a venerable ancestry in Early Modern philosophy. Descartes is the ancestor of dualists, Leibniz and Berkeley of idealists, Hobbes of reductive materialists, and Spinoza of neutral monists.[2] Illusionism, in contrast, is a relative newcomer, with no clear proponents before the 20th century. At least, this is the case in the West. In the Indian tradition, classical Yogācāra Buddhist thinkers, like Vasubandhu or Sthiramati, might have anticipated contemporary illusionism.[3] However, despite striking similarities,[4] this is a case of partial doctrinal convergence, not intellectual ancestry.

Western defenses of the eliminative claim appeared in the mid-20th century, but caricatural versions of the view had been considered and criticized earlier. In the early 1860s, Wilhelm Griesinger rejected the "dull and superficial materialism," which "throws overboard" the "general and valuable facts of the human consciousness" because these facts "cannot be grasped in the brain."[5] A literary incarnation of this 'dull' materialism is the nihilist Bazarov in Turgenev's *Fathers and Sons*.[6] In the 1920s, British philosopher C.D.

[2] Or property dualists, or anomalous monists. See Della Rocca (2008) for Spinoza as an anomalous monist.
[3] Garfield (2016); Chaturvedi (2023); Cermeño-Aínsa (2024).
[4] Including between my own view and those of Vasubandhu and Sthiramati (Chaturvedi 2023, 1517).
[5] Griesinger (1861, 6–7, my translation).
[6] Turgenev ([1862] 2009).

Broad rejected eliminativism about mentality in general, which he called "pure materialism."[7]

Behaviorism was introduced in the early 1900s as a methodological stance in psychology, and set aside consciousness as a scientific object of study. Its proponents did not explicitly defend the eliminative claim. They only suggested an "elimination of states of consciousness *as proper objects of investigation in themselves*."[8] Their claim was methodological:

> The time seems to have come when psychology must discard all reference to consciousness.[9]

Still, critics of behaviorism sometimes targeted interpretations of the view making it equivalent to eliminativism.[10] Watson himself made ambiguous claims, apparently going beyond methodology.[11] Whether later philosophical cousins of behaviorism (e.g., to be found in the work of Ryle, or the late Wittgenstein) came close to eliminativism is hard to assess. Still, I think these views were not eliminativist.

One must wait until the mid-20th century to see versions of the eliminative claim explicitly defended, in the work of Alf Ross, B.A. Farrell, W.O. Quine, Paul Feyerabend, and Richard Rorty.[12] The label "eliminative materialism" was then introduced by James Cornman, a critic of the view, in the 1960s.[13] By the 1980s, the view was widely discussed,[14] and had found prominent forceful defenders—most notably, Daniel Dennett,[15] who claimed that "contrary to what seems obvious at first blush, there simply are no qualia at all."[16]

[7] Broad (1925, chap. 14).
[8] Watson (1913, 177, my emphasis).
[9] Watson (1913, 163).
[10] Williams (1934).
[11] "The behaviorist finds no evidence for 'mental existences' or 'mental processes' of any kind" (Watson 1924, 2).
[12] Ross (1941); Farrell (1950); Feyerabend (1963); Quine (1960); Rorty (1965).
[13] Cornman (1968a; 1968b).
[14] Eliminativism was also developed about other mental entities, like propositional attitudes (Churchland 1981).
[15] Dennett (1979; 1988; 1991).
[16] Dennett (1988, 74).

B. Theoretical eliminativism and new-wave illusionism

The eliminativism of those years—typically, Dennett's—was combined with versions of the appearance claim, acknowledging that phenomenal consciousness seems to exist: these views are technically forms of illusionism. However, two things distinguish them from more recent forms of illusionism, which are closer to my own view.

First, 1980s eliminativists acknowledged that phenomenal consciousness seems to exist, but they typically thought its appearance results from faulty reasoning or cognitive mistakes. Georges Rey thought it is the product of mistaken projections.[17] Dennett believed it stems from reasoning mistakes: the "failure to understand [features] of [a] successful explanation," [18] and a fallacy he called the "user illusion."[19] For Dennett, we are under a user illusion when we judge that the schematic grasp we have of our own minds is fully accurate, like the naïve user of a computer who thinks icons on the desktop are more than schematic representations of complex computational structures, and are what literally populate the computer. Dennett saw the appearance of phenomenal consciousness as a product of "early education," and "in surprisingly large measures an artifact of our immersion in human culture."[20]

Second, and relatedly, these views defined the denied entity *theoretically*. Dennett denied the existence of *qualia*,[21] understood as ineffable, intrinsic, private, and immediately apprehensible properties of our mental states. *Qualia* or *phenomenal properties*, thus understood, are posits of sophisticated theories, not entities that appear real to ordinary people. The appearance of phenomenal consciousness, in this sense, is probably not widely shared.

Around the 2010s, a new variety of illusionism, coming both from philosophy and cognitive science, gained traction, through the work of Derk Pereboom, Nick Humphrey, and Michael Graziano.[22] In these views, the appearance of phenomenal consciousness is widely shared and compelling. It is akin to perceptual illusions: it is not the result of avoidable reasoning mistakes, and it persists even if we reason impeccably. It is not a cultural product that widely varies among humans and can be uprooted with

[17] Rey (1983; 1995).
[18] Dennett (1991, 455).
[19] Dennett (1991, 311–12).
[20] Dennett (1995a, 702).
[21] Dennett (1988, 47).
[22] Pereboom (2009; 2011); Graziano (2013); Humphrey (2011).

reeducation. *Introspection*, not flawed reasoning, generates the illusion. In the mid-2010s, Keith Frankish introduced the label "illusionism,"[23] as well as distinctions that considerably clarified discussions—more on this shortly.

We should not fetishize labels. On a liberal reading ("big tent" illusionism), both sorts of views—1980s eliminativism and 2010s new-wave illusionism—are illusionists. They accept the eliminative claim and a version of the appearance claim. Still, the label "theoretical eliminativism" fits the first views well.[24] In these views, the appearance of phenomenal consciousness results from cognitive mistakes, and phenomenality is defined theoretically; its appearance is probably not widely shared. On the other hand, the label "illusionism" is ideally suited to the second sort of views. These views do not define phenomenality theoretically, and they see the appearance of phenomenal consciousness as a widely shared and compelling illusion, akin to perceptual illusions.

C. Mistakist, perceptualist, and exceptionalist illusionism

As it views the appearance of phenomenal consciousness as a shared compelling illusion rather than a theoretical mistake, new-wave illusionism is sometimes found marginally more acceptable than theoretical eliminativism. Why? Most consciousness researchers are realists, who believe that the existence of phenomenal consciousness is *obvious*. It does not help if we claim that their entrenched opinion stems from avoidable reasoning mistakes, as theoretical eliminativists do. In contrast, new-wave illusionism grants that realists are not primarily bad reasoners. This might explain why new-wave illusionism gained traction and received, for instance, a mark of (conditional) appreciation from David Chalmers, maybe their most influential opponent:

> if I were a materialist, I would be an illusionist.[25]

Still, new-wave illusionism did not supplant theoretical eliminativism. For instance, Peter Carruthers recently defended "qualia irrealism," which he saw as a "close relative of illusionism": a variety of theoretical eliminativism, on

[23] Frankish (2016).
[24] Kammerer (2021, 848).
[25] Chalmers (2018, 9).

which the appearance of qualia is not a universal and spontaneous illusion, but results from "distinctive forms of reflective thinking" and "philosophical argument."[26] Besides, even big tent illusionism remains sociologically marginal.[27]

What sort of illusionism do I defend? Earlier, I defined "phenomenal consciousness" non-theoretically, and I presupposed that the appearance of phenomenal consciousness is shared and compelling. Therefore, I reject theoretical eliminativism. My view is closer to new-wave illusionism, as I think that the appearance of consciousness is a shared and compelling introspective illusion. However, I also think that the appearance of phenomenal consciousness is *uniquely* compelling. The illusion of phenomenal consciousness has unique features that make it much harder to overcome than other illusions. This distinguishes my view from current forms of new-wave illusionism, which, as I will show in Chapter 4, fail to acknowledge the unique potency of the illusion of phenomenality.

We can classify the main versions of illusionism by the type of erroneous process they primarily use to model the illusion of phenomenal consciousness. Theoretical eliminativists employ a "mistakist" approach: they treat the illusion of phenomenal consciousness like a reasoning mistake. New-wave illusionists typically employ a "perceptualist" approach: they use the model of perceptual illusions. My own view is relatively closer to the perceptualist than the mistakist approach, but it moves away from perceptualism by emphasizing the sui generis and unique character of the illusion of phenomenal consciousness. It takes an "exceptionalist" stance. Exceptionalist illusionism sees the illusion of phenomenal consciousness as unique—e.g., as uniquely compelling. These three labels (mistakism, perceptualism, exceptionalism) are rough, and many views lie somewhere in between. Still, they give a helpful overview.

2.3. Clarifying illusionism

Let us clarify illusionism, focusing first on its *commitments*. If illusionism is true, phenomenal experiences do not exist. What exists, then?

[26] Carruthers (2019, 14).

[27] In the last PhilPapers survey, 4.51% of the respondents chose "eliminativism" as a view of consciousness. This went down to 3.29% among philosophers of mind and up to 9.43% among philosophers of cognitive science.

Illusionists at least think that the *illusion* of phenomenal consciousness exists. Most also view many other things as real. I think that walls, chairs, bodies, jokes, revolutions, brains, neurons, molecules, electrons, etc., exist. Mental states exist. Non-phenomenal forms of consciousness exist: we are access-conscious and self-conscious. We are also *quasi*-phenomenally conscious—more on this shortly. Moreover, as explained in Chapter 1, the ordinary word "consciousness" typically expresses a multi-criterial concept mobilizing both phenomenal and non-phenomenal concepts. So, illusionists about phenomenal consciousness can also coherently claim—with the relevant caveats—that consciousness *tout court* exists, and that we are conscious. More on this in Chapter 7.

Does illusionism imply physicalism? No. It aligns well with physicalism, in both the strong and the weak sense, but it does not entail it. Not in the strong "structure-and-dynamics" sense: illusionism could be true even if non-structural, non-functional, and *non-phenomenal* quiddities exist. Not in the weak negative sense either: illusionism could be true even if fundamentally mental, non-physical, *and* non-phenomenal entities exist, e.g., fundamentally mental but non-phenomenal propositional attitudes.

Two distinctions, due to Keith Frankish, clarify illusionism's commitments. The first is between *phenomenal* and *quasi-phenomenal* consciousness.[28] Suppose phenomenal consciousness does not exist. Nothing satisfies our phenomenal concepts. Still, deployments of phenomenal concepts (e.g., in introspection) could track real things. Whenever I introspect phenomenal pain, a certain type of *real, non-phenomenal state*—e.g., a brain state—arguably triggered my introspective deployment of the phenomenal concept of pain. This type of state might be reliably *tracked* by my introspective deployments of the concept. Following Frankish, call states like this *quasi-phenomenal* experiences of pain.

Phenomenal introspection deploys *reliable misrepresentations* of quasi-phenomenal experiences.[29] Quasi-phenomenal experiences have quasi-phenomenal properties, in virtue of which they are quasi-phenomenal experiences of certain types. The overarching phenomenon is quasi-phenomenal consciousness. Quasi-phenomenal states could be brain states. They could also include environmental or bodily features. They might be best described at the physiological level, or the computational level, etc.

[28] Frankish (2016, 15–16).
[29] For the concept of reliable misrepresentation, see Mendelovici (2013).

They might or might not form a single salient natural kind. What matters is that these states, which trigger phenomenal introspection and which our introspective deployments of phenomenal concepts track, almost certainly exist. Still, they are not phenomenal. They do not satisfy the characterization borne by our phenomenal concepts. They are reliably tracked, but *mischaracterized* by these concepts. There is *nothing it is like*, in the sense of the expression corresponding to the substantive grasp offered by our phenomenal concepts, to have quasi-phenomenal experiences—of red, of pain, etc. We are quasi-phenomenally conscious, but not phenomenally conscious.

One might wonder: I talked of quasi-phenomenal consciousness, but why use such a cumbersome name? Illusionists think this form of consciousness exists: Why not call it by what it *is*, rather than what it is not? For now, this would be premature. Illusionism is silent on the nature of quasi-phenomenal consciousness. Identifying it falls on science, not philosophy. Still, could we not call quasi-phenomenal consciousness something less clumsy, like "experiential consciousness," "awareness," etc.? The problem is that these convenient ordinary terms are also sometimes used to refer to *phenomenal* consciousness itself. Denying phenomenal consciousness but not experiential consciousness opens the door to misunderstandings. It is preferable to refer to this form of consciousness *via* an expression that makes it clear that we grasp it by misrepresenting it. "Quasi-phenomenal consciousness" does the job. Other expressions would work too ("phenomenalish consciousness," "pseudo-phenomenal consciousness," or "merely apparent phenomenal consciousness"), but they are even clumsier.

What I called "illusionism" so far is *strong* illusionism. Frankish, in his second key distinction, opposes it to *weak* illusionism.[30] Strong illusionism states that phenomenal consciousness does not exist, but seems to exist. Weak illusionism states that it exists, but does not have certain features it seems to have.

In a liberal sense, almost all consciousness researchers are weak illusionists. They admit that phenomenal consciousness is not entirely as it seems to be. Someone who thinks that visual phenomenal experiences seem fully detailed across the visual field, although they are not, would count as a weak illusionist. However, even staunch realists find this plausible.[31]

[30] Frankish (2016, 15–16).

[31] Chalmers (2018, 52). Frankish (2023) recently claimed the weak vs. strong illusionism distinction is degreed.

Therefore, I reserve the weak illusionist label for views in which phenomenal consciousness lacks some *central* features it (more or less) *compellingly* seems to have.

Weak illusionists are phenomenal realists. They can be called "weak realists," by contrast with *strong* phenomenal realists, who think that phenomenal consciousness has the central features it (more or less) compellingly seems to have. In what follows, "illusionism," unqualified, means "strong illusionism."

The landscape of positions on consciousness is complex, but broad categories are helpful. In what follows, I will contrast illusionism with two paradigmatic alternatives. On the one hand, there are *strong realist* views of phenomenal consciousness. In these views, phenomenal consciousness exists, and it really has the central features it seems to have. Typically, this means that it is introspectively obvious and non-physical; metaphysically, these views are *anti-physicalist* (dualist, idealist, or neutral monist). On the other hand, there are *weak realist* views, which are also weak illusionist views. In these views, phenomenal consciousness lacks some of the central features it seems to have. Typically, the denied features are the mysterious ones I described earlier; metaphysically, these views are *physicalist*. These two camps are the two most popular alternatives to illusionism. It is useful to keep them in mind when discussing possible answers to the illusionist arguments.[32]

Finally, one last category is useful. Imagine a weak illusionist who claims that phenomenal consciousness lacks *a lot* of the central features it seems to have—notably, that it lacks *all* the key mysterious features corresponding to phenomenal concepts' characterizations. For this weak illusionist, phenomenal consciousness's real features are only those had by what the strong illusionist calls "quasi-phenomenal consciousness." This view takes weak illusionism to the extreme. It does not differ substantively from strong illusionism, but only verbally so. What this sort of weak illusionist calls "phenomenal consciousness" is exactly what the strong illusionist calls "quasi-phenomenal consciousness." What the strong illusionist calls "phenomenal consciousness" can be called "phenomenal consciousness as it seems to be" by this sort of weak illusionist. Call this extreme version of weak illusionism "verbal weak illusionism."

[32] See also Frankish (2016, 13–14) on radical vs. conservative realism about consciousness.

2.4. Illusionism's two challenges

A. Supporting illusionism

Illusionists face two main challenges. First, they must *argue* for the view. This is no trivial task. Since the view includes the eliminative claim, it is very counterintuitive.

Views on which the appearance of phenomenal consciousness is a shared and compelling illusion might be more plausible than theoretical eliminativism, and more palatable to those with strong realist tendencies, since they grant that realists do not need to be bad reasoners. Still, they take an extra dialectical risk. If phenomenality is not defined theoretically, and if its appearance is shared and compelling, phenomenal consciousness is something that seems to exist to ordinary people: the eliminative claim is revisionary. By contrast, if phenomenality is a theoretical posit, whose appearance results from flawed reasoning, eliminativism is not revisionary: it contradicts sophisticated philosophical views, not ordinary people's views. Plausibly, revisionary views require more evidence than others.

Take the measure of the challenge. The eliminative claim is not *just* counterintuitive and revisionary. Many find it completely unbelievable. John Searle saw it as "an obvious and self-refuting falsehood" and came close to denying the sanity of its defenders.[33] Galen Strawson called it "the silliest claim ever made."[34] Chalmers, despite the conditional praise mentioned earlier, stated that it is "obviously false,"[35] that it "denies the evidence of our own experience," and stressed that "this is the sort of thing that can only be done by a philosopher."[36] Others called it "crazy,"[37] "impossible" and "absurd,"[38] "self-defeating," and "incoherent."[39]

I understand this sentiment. Illusionism stands at the limit of what we can understand and accept. We should not sweep its deeply counterintuitive character nor its apparently unintelligible nature under the rug. My version of illusionism is exceptionalist: it grants that the illusion of phenomenal consciousness is compelling, and even uniquely compelling. This concession is

[33] Searle (1997, 120–22).
[34] Strawson (2018).
[35] Chalmers (2018, 54).
[36] Chalmers (1996, 188).
[37] Strawson (1994, 101); Frances (2008, 241).
[38] Nida-Rümelin (2016, 163, 170).
[39] Goff (2016, 84–85).

necessary for the view to be ultimately convincing. I even think that, to grasp illusionism as I see it, we *must* appreciate its strangeness, its counterintuitive character, and its apparently unintelligible nature. If you find illusionism intuitive or trivial, you do not understand the view like I do.

So, the first challenge for illusionists is to argue for their view, and notably for the eliminative claim. To meet this challenge, in Chapter 3, I deploy two arguments working in tandem. In Chapter 7, I examine objections to illusionism, most of them directed at the eliminative claim.

B. The illusion problem

Illusionists face a second key challenge. If you think that phenomenal consciousness appears to exist, you must explain *why*. You have to solve the *appearance problem*. Illusionists, and virtually all realists, accept the appearance claim. They have to solve this problem.

Illusionists and realists face different versions of the problem. Realists have to explain why phenomenal consciousness appears to exist and why this appearance conduces to knowledge. They specifically face the *knowledge problem*. Illusionists have to explain why it seems to exist even though it does not exist. They specifically face what Frankish calls the *illusion problem*.[40]

Solving the illusion problem is one of the main tasks of illusionists. It occupies the central chapters of this book. In Chapter 4, I show that solving the illusion problem notably requires explaining our mystery intuitions. Therefore, the illusion problem is closely related to what Chalmers called the "meta-problem of consciousness."[41] The meta-problem of consciousness is, roughly, a problem about the existence of the hard problem: Why do we think there is a hard problem? More generally, it is the problem of explaining mystery intuitions.

Finally, in my exceptionalist form of illusionism, the illusion of phenomenal consciousness is uniquely compelling. Explaining why requires understanding why illusionism itself is so hard to accept and represent to ourselves: this is what I call the *illusion meta-problem*.[42] I see the illusion meta-problem as a *subpart* of both the illusion problem and the meta-problem of

[40] Frankish (2016, 37).
[41] Chalmers (2018).
[42] Kammerer (2018a).

consciousness. Of course, some people tackle the illusion problem and the meta-problem of consciousness, but deny that the illusion of phenomenal consciousness is uniquely compelling, and therefore, that this illusion meta-problem even arises in the first place.

We have a clearer idea of what illusionism is. I will now give a brief argument for the appearance claim, and lay the foundations for the arguments for the eliminative claim deployed in the next chapter.

2.5. The appearance claim

The appearance claim says that phenomenal consciousness seems to exist. More precisely, in my view, introspection gives us shared and compelling dispositions to believe that we have phenomenal experiences. These dispositions are compelling. First, they are *robust*: conditional on the use of introspection, they are hard to destroy and are cognitively impenetrable. Second, they are *strong*, and notably stronger than perception-based dispositions to form beliefs about our environment. This is shown by the fact that it is easier to admit that perception, rather than phenomenal introspection, is illusory. Still, these dispositions are not overwhelmingly strong. We *can* believe in illusionism.

Unlike mystery intuitions, the appearance of phenomenal consciousness does not have a conceptual origin. Concepts of phenomenal properties do *not* characterize them as instantiated. Even strong realists concede that we can think of some phenomenal property as never instantiated: e.g., we can conceive of a peculiar super-intense pleasure felt on the tip of the nose, which never occurred anywhere. The primary source of the appearance of phenomenal consciousness is not conceptual reflection, but *introspection*.

The appearance claim is not very controversial. Most thinkers *reject* the eliminative claim and think that phenomenal consciousness is obviously real. A fortiori, they are committed to something like the appearance claim. Therefore, I will only briefly argue for the claim. Some objections to it are examined in Chapter 7.

So, why do I believe the appearance claim? First, because of what happens in my own case. When I introspect, I strongly tend to believe that I have phenomenal experiences. Contrast this with perception. Suppose I look at a white wall: I am perceptually disposed to believe there is a white wall in

front of me, and introspectively disposed to believe I have a phenomenal experience of a white wall. Both dispositions are robust and cognitively impenetrable. The first, however, is relatively easy to resist. You might convince me that I face cleverly arranged mirrors creating the illusion of a wall. The second disposition is much stronger. Outside of deep philosophical reflection, nothing can override it.

So, in my case, the appearance of phenomenal experiences seems introspective and highly—if not uniquely—compelling. Besides, many others also claim to have phenomenal experiences. They tell me that they know about them by focusing inward. It is extremely hard to convince them that they do not really have them. This suggests that this introspective and compelling appearance of phenomenal consciousness is *widely shared*.

So, in introspection, phenomenal consciousness *compellingly* seems to us to exist. Here, an issue arises. Earlier, I defined "phenomenal consciousness" by deferring to the grasp offered by phenomenal concepts. Now, suppose others are disposed to believe that they enter certain states they call "phenomenal experiences." Still, how can I establish that the concepts they use to think about these states are like my phenomenal concepts, so that what appears to them to exist is what *I* call "phenomenal experiences"? I should establish this to claim that the appearance of phenomenal consciousness is widely shared.

Appealing to psychological similarity helps: plausibly, other humans are like me. Still, this similarity might not hold across the board. One could add that, plausibly, tendencies to believe that are strong and robust at the individual level (like mine regarding phenomenal experiences) are often widely shared, while tendencies that vary across individuals are often weaker and less robust. Still, even this is not always true: religious upbringing creates strong, robust, but inter-individually variable tendencies to believe.

Here is a solution. Phenomenal consciousness, I claimed, is conceptually distinct from access-consciousness and self-consciousness. Besides, mystery intuitions are conceptual in origin: our phenomenal concepts implicitly characterize their referents as obvious, irreducible, and undetectable.

Suppose all of this is right in my case. This helps identify phenomenal concepts in others. Suppose others introspectively tend to believe they enter certain states. If they are able to conceptually distinguish these states from access-conscious and self-conscious states, and if they tend to view them as obvious, irreducible, and undetectable, this makes it quite likely that they

represent these states with phenomenal concepts, and that phenomenal experiences appear to them to exist.

In the previous chapter, I gave reasons to think this happens. Therefore, we have shared and compelling introspective tendencies to believe in phenomenal consciousness. The appearance claim is true.

2.6. Approaching the eliminative claim

Given the eliminative claim, phenomenal consciousness does not exist. There are no phenomenal experiences. Nothing real satisfies the characterization borne by our phenomenal concepts.

I do not define "exist." I suppose it expresses a primitive concept we all grasp. We understand what it is to exist for things that exist—tables, electrons, galaxies, revolutions, jokes, etc. We understand what it is to fail to exist for things that fail to exist—Santa Claus, phlogiston, witches, ghosts, sirens, the divine right of kings, etc. This is enough to make sense of the eliminative claim.

I treat "existing," "being real," "being a part of reality," and "being in the world" as synonymous. Sometimes, we distinguish those. Numbers or properties (as universals) might be real without existing and without being in the world. However, if phenomenal experiences are real, they are not properties, but *instances* of phenomenal properties. They exist, and they are in the world. We can ignore debates about the metaphysics of abstracta. I also set aside whether phenomenal properties *qua* properties are real and subsist in a Platonic realm. What matters is whether they are instantiated.

The eliminative claim is not a priori true. Our concept of phenomenal consciousness is coherent. Pure reasoning is not enough to show that it is not satisfied. The eliminative claim is not a priori false either. The debate between eliminativism and realism is an a posteriori one.

Should the realist accept this framing of the debate? I think so. Many think that phenomenality is real but concede that the idea of a world without phenomenal states is coherent. Maybe the early universe was like this. Zombie worlds are like this. So, if you think that phenomenality exists, it is not via reasoning alone, but because of *empirical data* you have about our world. Here, "empirical data" has a broad meaning. It excludes the output of pure reasoning, but includes introspective data about what happens in our minds.

One might retort that, while the idea of a world without phenomenal experiences is coherent, *that we are* in such a world is a priori incoherent and can be ruled out independently of empirical data: merely *thinking* about phenomenal consciousness puts us in a position to know a priori that it exists. This stance is not widespread, but it deserves consideration.[43] One can contend that phenomenal concepts rationally require their users to judge that at least *some* phenomenal concept is satisfied, even if this cannot always be the exact concepts used, since undeniably we can think that some phenomenal properties are never instantiated (e.g., the super-intense pleasure felt on the tip of the nose mentioned earlier).

Similarly, while the idea of a world in which I do not exist is coherent, there is arguably something incoherent in the view that *I* (thought with an indexical) actually do not exist.[44] This incoherence does not reduce to the self-verifying nature of the thought, since some self-verifying thoughts can be coherently taken to be false by the thinker; suffering from amnesia, I can coherently think that François Kammerer does not exist. This incoherence signals deeper conceptual facts, by which a concept exerts a rational constraint on the thinker—say, making it a priori that the concept *I* is satisfied.

This objection leads to the following dialectic. Suppose phenomenal concepts exert this rational constraint on their users. I then ask whether *we* use phenomenal concepts, exerting such rational constraint, or closely related *phenomenal** concepts—functionally similar concepts without the constraint. This second question seems a posteriori. The objector might retort that it is in fact also an a priori matter, since we use phenomenal concepts, which might also rationally require their user to think they use phenomenal concepts and not phenomenal* concepts. I then ask whether we use concepts exerting this other rational constraint, or other functionally similar concepts without this constraint, and claim that this is an a posteriori question, etc. This leads to an uninteresting stalemate. A reason to leave the stalemate via the a posteriori exit is that claims that a truth is a priori often only help debates when they appeal to rational constraints recognized by all parties of a debate, which does not seem to be the case here. We are better off treating the question as an a posteriori one, at least for now. In Chapter 7, I explain why eliminativism might *seem* a priori false, even if it is a posteriori true.

[43] Nida-Rümelin (2016) and Balog (2023) might come close to this line of reasoning.

[44] This is related to rationalist interpretations of the Cartesian cogito, opposed to empiricist ones (Billon 2015).

2.7. Data for and against the eliminative claim

The eliminative claim is an a posteriori truth or an a posteriori falsity. To justify it or its negation, we need empirical data. "Data" means propositions we can legitimately treat as given at a time. "Empirical" is construed broadly, to include introspective data.

Data for and against the eliminative claim are divided along two lines. First, distinguish *tendentious* from *non-tendentious* data.[45] Non-tendentious data is data that both eliminativists and phenomenal realists can accept prima facie. Tendentious data is prima facie unacceptable by at least one camp. Second, distinguish *direct* and *indirect* data. Direct data is data from which we can deduce eliminativism or realism. Indirect data only supports one of the two via non-deductive inferences, like inferences to the best explanation, which support views required by the best explanations of our data. Call direct data for realism "phenomenal data."

I do not endorse a particular view of explanation, but I presuppose that explanations require deducing the *explanandum* from the *explanans*,[46] thus making the *explanandum* intelligible.[47] I also suppose that explanations can have various explanatory virtues aside from the fact that they make it possible to deduce their explanandum: simplicity, elegance, internal coherence, external coherence (i.e., coherence with other accepted views), and the capacity to solve or avoid problems.

Eliminativists and realists disagree about the phenomenal, but they agree on other things: on features of trees, electrons, clouds, galaxies, bacteria, quarks, jokes, etc. They agree on many properties of human brains and behaviors. Generally speaking, they agree on physical data, formulated with phenomenal concepts: physical data is non-tendentious.[48] Since physical and phenomenal concepts are conceptually independent, we cannot deduce the existence or non-existence of the phenomenal from physical data. Physical data, here, can only serve as indirect data. Some, like analytic functionalists or logical behaviorists, deny this independence and think that

[45] For the distinction, see Levine (2001, 144), who borrows it from Rey (1997).

[46] Or at least to deduce a probability distribution, over alternative ones, among which the *explanandum* is to be found. I set aside this precaution from now on.

[47] For the view that prominent theories of explanation (whether epistemic or ontic) imply that good explanations allow to deduce the *explanandum* from the *explanans*, see Levine (2001, 70–76).

[48] Even *idealists* agree with others about physical data concerning structural and causal features. They just insist that these features are ultimately grounded in the mental.

we can deduce the existence of the phenomenal from physical data, but I assume that they are wrong.

We do not have direct data *for* eliminativism, but my opponents often think that we have direct data for realism—phenomenal data. Typically, they think that this data is *introspective*. Eliminativists reject such introspective phenomenal data. Introspective phenomenal data is *tendentious*.

Is non-tendentious phenomenal data impossible by definition? No. In principle, both camps could first agree on some data. Further reflection could then show that the existence of phenomenal experiences is deducible from this data—e.g., using non-obvious paths and appealing to implicit conceptual truths only established via thorough conceptual analysis. Eliminativists might then *react* by rejecting the data, but it would remain non-tendentious in my definition. For some, *normative data* about epistemic justification, well-being, or moral status is non-tendentious phenomenal data in this sense: using only logical and conceptual truths, we can deduce the existence of the phenomenal from such data, even if this deduction is not obvious. For now, however, I suppose that all phenomenal data is introspective and tendentious. I discuss normative data in Chapters 9 and 10.

To summarize, we have abundant non-tendentious physical data, which can only serve as indirect data in the debate between eliminativism and realism. Realists claim that we also have introspective phenomenal data. Such data is tendentious. Other potentially relevant data is set aside for now.

* * *

I presented illusionism about phenomenal consciousness, and I clarified the view. I gave a brief argument for the appearance claim. Then, I approached the eliminative claim. I will now argue for the eliminative claim.

3
Arguing for Illusionism

To be a good philosopher you have to be dry, clear, and without illusion
—Stendhal

The eliminative claim seems outrageous to many. Arguing for it is the first major challenge faced by illusionists. In support of the eliminative claim, I present two arguments working in tandem. First, the *Argument from Mysteries*, concluding that, if we have good reasons to reject phenomenal data, we should believe eliminativism. Second, the *Argument from Irrelevance*, concluding that we have good reasons to reject phenomenal data.

3.1. The arguments in a nutshell

Here is the overall argumentative strategy. Phenomenal consciousness seems *mysterious*. These apparent mysteries put pressure on phenomenal realism. Do we have convincing direct data entailing that this mysterious entity exists? That is, do we have convincing phenomenal data? Or, maybe, does positing this mysterious entity help us explain our non-phenomenal data? If we answer "no" to both questions, we should deny the existence of phenomenal consciousness. A mysterious entity, whose existence is not directly supported by convincing data, and which does not help us explain other data, is not something we should believe in.

Does positing phenomenal consciousness help us explain other things? No. Considering our best sciences and reflecting on plausible metaphysical assumptions, we see that we can explain the non-phenomenal as well if we deny the phenomenal as if we posit it. We reach an intermediate conclusion: since phenomenal consciousness seems mysterious, and since positing the phenomenal does not help us explain other things, we should only think phenomenal consciousness exists if we have convincing

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0004

phenomenal data. If we have good reasons to reject phenomenal data, we should believe the phenomenal does not exist. This is the *Argument from Mysteries*.

We have good reasons to discard putative phenomenal data. Why? Among the non-phenomenal entities about which we have data, some processes—physically conceived *phenomenal judgments*—should carry our putative phenomenal data: *if* we have phenomenal data, these phenomenal judgments carry it. A consequence is that we should explain these judgments better by positing the phenomenal than by denying it. However, positing the phenomenal does not help us better explain phenomenal judgments. This gives us a good reason to reject phenomenal data. This is the *Argument from Irrelevance*.

So, phenomenal consciousness seems mysterious; it does not help explain anything else, and we have good reasons to reject direct data for its existence. Therefore, we should deny its existence.

Philosophical arguments are rarely decisive. My arguments make no exception. They are only the first building block of the case for the eliminative claim, which must ultimately be assessed holistically.

3.2. The Argument from Mysteries

Here is an outline of the first argument:

Premise 1 (*Mystery*): Phenomenal consciousness seems mysterious.
Premise 2 (*Explanation*): Phenomenal realism does not account for our non-phenomenal data better than eliminativism.
Premise 3 (*Razor*): If phenomenal consciousness seems mysterious and phenomenal realism does not account better for our non-phenomenal data than eliminativism, then, if we have good reasons to reject phenomenal data, we should believe eliminativism.
Conclusion (*Conditional Eliminativism*): If we have good reasons to reject phenomenal data, we should believe eliminativism.

Here, "eliminativism" and "phenomenal realism" have a broader sense than before. They refer to our *best overall views of reality* including the eliminative claim or its negation.

The argument is valid. Let us examine its premises.

A. Premise 1—Mystery

Mystery states that phenomenal consciousness seems mysterious. I read it as reiterating some presuppositions of Chapter 1: we have shared compelling and conceptual mystery intuitions about phenomenal consciousness—e.g., regarding its obviousness, irreducibility, and undetectability.

At this stage, I assume *Mystery*. Some deny it. Those type-A materialists who think that phenomenal concepts are behavioral or functional concepts deny it. Many weak realists deny that our mystery intuitions are compelling and conceptual. I examine objections to *Mystery* later, in Chapter 7. Below, when discussing *Razor*, I also show that weaker versions of *Mystery* might be sufficient to power the argument.

B. Premise 2—Explanation

Explanation states that phenomenal realism does not account for our non-phenomenal data better than eliminativism.

Non-phenomenal data is data from which we cannot deduce phenomenal realism. I assume, against analytic functionalism or logical behaviorism, that it comprises *physical* data. This includes data provided by physics, chemistry, and biology; much of the data provided by behavioral, psychological, and social sciences (when it can be formulated functionally); much ordinary data about tables, rocks, smiles, etc. Roughly, physical data is data we can have about a zombie world.

Is there non-phenomenal data beyond physical data? *Normative* data is a plausible candidate. I postpone discussions of normative data to Chapters 9 and 10. I assume, for now, that non-phenomenal data is physical data.

Why think that phenomenal realism does not explain physical data better than eliminativism? I see two reasons to believe so.

First, consider the *commitments* of our best accounts of physical data. None of our best scientific theories appeals to phenomenal consciousness. They posit merely physical entities, grasped with physical concepts. Aristotelian physics appealed to 'souls' to explain natural processes, but it is a cliché in the history of science that Galileo and Descartes de-mentalized physics. Our current physical theories do not posit anything mental in their explanations—a *fortiori* anything phenomenal. The same applies to chemistry and biology. Quantum physics employs the concept of an observer,

but it is usually understood non-phenomenally. In some views,[1] phenomenal consciousness plays a role at the quantum level, but these views are not standard, and it is not sure that they assign a *key* role to phenomenal consciousness, which could not be played by a non-phenomenal equivalent—see, below, the idea of *functional substitutes*.

Therefore, our best explanations of physical data posit physical entities, described with physical concepts. Since physical and phenomenal concepts are conceptually independent, we can conjoin these explanations with the eliminative claim without changing how they account for the data. Eliminativism explains physical data as well as realism.

Second, consider the *causal closure* of the physical. The causal closure (or "completeness") of the physical is a popular principle, often judged plausible. In a modest version, it states that every physical effect has a sufficient physical cause.[2] This strongly suggests that every bit of our physical data is deducible from some other physical truths. Now, compare eliminativist and realist explanations of physical data. There is no strong reason to think that realism fares better than eliminativism when it comes to virtues such as simplicity, elegance, and internal coherence. In fact, eliminativism is probably simpler and more elegant than realism, but my argument does not dwell on this. We can also set aside virtues of external coherence, and problem-avoiding or problem-solving: I think that realism is inferior to eliminativism in these respects, but this is what motivates *Razor* and I ignore this point here.

What remains to compare realist and eliminativist explanations of physical data? Their capacity to *deduce* physical data. Here, the causal closure of the physical becomes relevant. It implies that all physical data can be deduced from some other physical truths—typically, about temporally prior phenomena. Besides, since physical and phenomenal concepts are conceptually independent, any physical truths from which we deduce physical data can be conjoined with the eliminative claim without impacting the deduction. Therefore, *some* eliminativist views can deduce all our physical data. This suggests that realism does not fare better in this respect either. Therefore, it does not overall better explain physical data.

Objections can be raised to these two lines of reasoning. Against the argument appealing to the commitments of our best explanations of physical data, one could stress that, outside of the natural sciences, we often

[1] Hameroff and Penrose (2014); Chalmers and McQueen (2022).
[2] Papineau (1993a).

appeal to phenomenal experiences to account for physical data. Cognitive scientists try to explain behavior by appealing only to physical processes, but they still frequently mention phenomenal experiences as explanatory factors. Social scientists explain actions by mentioning people's feelings and experiences: e.g., historians explain Hitler's rise to power by his exploitation of feelings of humiliation in post-WWI Germany.

Does this mean realism explains physical data better than eliminativism? No. Suppose cognitive and social scientists appeal to phenomenal experiences to explain some physical data. To do so, they must take these phenomenal experiences to ground certain causal powers, making them able, for instance, to cause behavior. Since phenomenal concepts do *not* characterize their referents as having causal powers, if scientists think phenomenal experiences have causal powers, they must admit, at least implicitly, principles specifying these powers. If, in fact, phenomenal concepts characterize their referents as causally powerful, then these principles are just built into our phenomenal concepts.

We can then form concepts of *functional substitutes* of these powerful experiences: physical, *non-phenomenal* concepts of entities bearing the same causal powers as these experiences. Positing such functional substitutes, eliminativists explain the same physical data as realists: one can indifferently claim a given phenomenal experience is present or absent (with its functional substitute being present) and provide the same explanations.

This claim is not original, and it has been clear at least since the early behaviorists:

> one can assume either the presence or the absence of consciousness [. . .] without affecting the problems of behavior by one jot or one tittle.[3]

Without a reason to think that explanations of physical data appealing to functional substitutes are inferior to those appealing to phenomenal experiences, we can maintain that eliminativism explains physical data as well as realism. Of course, logical behaviorists or analytical functionalists, for whom phenomenal concepts *are* indistinguishable from concepts of functional substitutes, will deny *Explanation*, but I presupposed that these views are wrong.

[3] Watson (1913).

Against the argument mobilizing the causal closure of the physical, interactionist dualists deny causal closure. Still, to justify the explanatory superiority of realism, they must suppose that some physical data cannot be explained (or not as well) by positing only physical entities: our best explanations of the physical must require positing phenomenal experiences *qua* phenomenal, and not *qua* grounding physical properties. This implies that these explanations cannot be given with functional substitutes. If these explanations correspond to those given in cognitive and social science, this also answers the first argument.

Some speculative interactionist dualist views have this consequence. For instance,[4] Brad Saad speculates that phenomenal states, in virtue of fundamental teleological laws, tend to cause the effects they rationalize: e.g., they tend to cause correct judgments about them, which they rationalize by making them correct. Plausibly, such a teleological-rationalizing explanation of our judgments about phenomenal states could *not* be given merely with concepts of functional substitutes, since functional substitutes would not rationalize these judgments. If these teleological laws hold, appealing to phenomenal states allows for a uniquely superior account of physical data. However, the existence of these laws is extremely speculative. Interactionist dualism in general is widely rejected, so that this sort of criticism of *Explanation* is unlikely to be dialectically effective.

One could also concede causal closure, but contend that eliminativists cannot deduce *all* physical data, as they deny some behavioral and functional features needed for the deduction. This happens if logical behaviorists or analytical functionalists are correct and phenomenal concepts are behavioral or functional concepts, but I assume that they are wrong. Finally, one could concede causal closure but insist that, while eliminativists can deduce all physical data, realists do it better, as phenomenal explanations are superior—e.g., more illuminating. (This might require using a longer list of explanatory virtues than the one I used.) The 'phenomenal powers' view plausibly entails this.[5] On this view, which her main defender, Hedda Hassel Mørch, sees as naturally combined with panpsychism, phenomenal properties metaphysically necessitate their effects (i.e., they have non-Humean causal powers) in an intelligible way.[6] For instance, *feeling pain* intelligibly causes *efforts to*

[4] For other views affording similar defenses, see Chalmers (2018, 41) on "meta-problem nihilism."
[5] Mørch (2020). She ascribes a version of the view to Langsam (2011).
[6] Mørch (2020, 133).

make the pain stop. The physical data explained by phenomenal experiences can also be explained by functional substitutes, but the resulting explanation is not as illuminating.

This answer faces problems. First, it appeals to highly speculative views, such as the phenomenal powers view, which, in its standard panpsychist versions, leads to counterintuitive consequences and deep problems, like the so-called combination problem."[7] Second, this answer rejects the conceptual independence of phenomenal concepts and physical concepts and the impossibility of a priori deducing causal propositions from phenomenal propositions, which is costly. Third, even conceding that it is intuitive that some phenomenal properties necessitate some effects, it does not seem intuitive for physical, non-phenomenal effects, which limits the attractiveness of the view. Consider the idea that phenomenal pain intelligibly necessitates an effort to make the pain stop. It might be intuitive if *effort* is understood as a phenomenal experience, but not if understood as a process about which we only have physical data (e.g., a pattern of neural firing).[8] Finally, our intuition that phenomenal experiences intelligibly necessitate some effects might occur even without genuine intelligible necessitation. A *false impression* of necessitation, stemming from our *humean, too humean* habituation to apparent transitions between phenomenal experiences and other experiences or behaviors, could be all that happens.[9]

So, eliminativism explains physical data at least as well as realism. Supposing here that non-phenomenal data is physical, we reach *Explanation*: phenomenal realism does not account for our non-phenomenal data better than eliminativism.

C. Premise 3—Razor

Razor states that if phenomenal consciousness seems mysterious and phenomenal realism does not account better for our non-phenomenal data than eliminativism, then, if we have good reasons to reject phenomenal data, we should believe eliminativism.

Why believe *Razor*? Suppose that eliminativism fares as well as realism when it comes to deducing our data, simplicity, elegance, and internal

[7] Chalmers (2016); Roelofs (2019).
[8] Something Mørch (2020, 137) concedes.
[9] But see Mørch (2017) for a discussion of this objection.

coherence. Since phenomenal consciousness seems mysterious, accepting its existence creates tensions with other attractive views and generates corresponding theoretical problems. Hence, realism is inferior to eliminativism regarding external coherence and the capacity to solve or avoid problems, while both fare similarly otherwise. An inference to the best explanation favors eliminativism.

I assume here that eliminativism indeed fares at least as well as realism regarding simplicity, elegance, and internal coherence. So, if it does as well as realism when it comes to deducing all of our data, and if phenomenal consciousness seems mysterious, we should believe eliminativism.

The key idea is that, as phenomenal consciousness seems mysterious, realism creates tensions with attractive views, as well as theoretical problems. Phenomenal consciousness seems irreducible to anything physical. Therefore, realism is in tension with physicalism, an independently attractive metaphysical thesis—parsimonious, elegant, and congruent with the success of natural sciences. Phenomenal consciousness seems epistemically irreducible and impossible to explain scientifically. So, realism is in tension with the attractive view that our best sciences' methods are in principle sufficient to explain mental phenomena. Our intuition of undetectability renders phenomenal consciousness at odds with attractive principles of scientific inquiry, such as the principle that reality is third-personally accessible. Our intuition of obviousness makes it at odds with the attractive naturalistic view that our knowledge relies on natural and fallible representational processes. Overall, mystery intuitions make phenomenal consciousness seem *anomalous*. Admitting anomalous entities conflicts with attractive views regarding the simplicity and uniformity of nature.[10] Since mystery intuitions are conceptual, these issues persist as long as we take phenomenal concepts to be satisfied.

We can make the same point by stressing the theoretical gain eliminativism generates. Given eliminativism, the hard problem evaporates and the explanatory gap disappears: if phenomenal consciousness does not exist, it does not have to be explained. Eliminativists easily defend physicalism against consciousness-related arguments. Take Frank Jackson's Mary. Illusionists deny that, upon seeing a red object, she acquires substantive new knowledge bearing on a new fact irreducible to those she already knew. They grant

[10] For an illusionist argument directly appealing to this anomalousness, see Frankish (2016, 27–28).

that she acquires an entirely new belief, but it is only a false belief about the instantiation of phenomenal properties.[11] She acquires no new knowledge about anything non-physical, and her case does not threaten physicalism. Take the zombie argument. Illusionists grant that zombies are conceivable and possible. They simply add that they are actual, since we do not differ from zombies.[12] The possibility of zombies does not show anything about our world having non-physical features. Finally, illusionism can dissolve various problems in consciousness science—more in Chapter 8.

These theoretical gains are considerable. Some, like Dennett, think that they are enough to make illusionism the "obvious default theory of consciousness": alternatives to illusionism accept a mysterious form of consciousness and appeal to "magical hypotheses."[13] Michael Graziano concurs:

> there are now only two broad classes of explanation for consciousness: magic and illusionism.[14]

I, for one, do not think that these theoretical gains are decisive. My assumption is more modest: everything being equal, they make eliminativism preferable. Granting this does not imply that we should deny *all* seemingly mysterious entities.

So, if phenomenal consciousness seems mysterious, and if eliminativism explains our data as well as realism, we should believe eliminativism. Is this still plausible if we understand "seeming mysterious" more weakly, in a way aligned with a weaker version of *Mystery*? Maybe. The apparent mysteries of phenomenal consciousness do not have to be all those I assumed. A mere appearance of epistemic irreducibility could suffice. Still, mystery intuitions probably need to be conceptual; otherwise, the problems faced by realists would not simply depend on them being realists, but on other factors as well, and this would threaten the preferability of eliminativism.

Let us summarize. If phenomenal consciousness is mysterious, and if eliminativism explains all of our data as well as realism, eliminativism is preferable to realism. Now, our data is either phenomenal or non-phenomenal. Suppose eliminativism accounts for non-phenomenal data as well as realism. We are left with phenomenal data which, by definition, cannot be accounted

[11] Pereboom (2011, 25).
[12] Dennett (1991, 406)
[13] Dennett (2016, 65)
[14] Graziano (2024, 12).

for by eliminativism. So, if eliminativism explains non-phenomenal data as well as realism, we should admit that eliminativism explains all of our data as well as realism if (and only if) we should reject putative phenomenal data. Should we reject phenomenal data? We should do it if (and only if) we have good reasons to do so. (More on what counts here as *good reasons* later).

We reach *Razor*: if phenomenal consciousness seems mysterious *and* phenomenal realism does not account better for our non-phenomenal data than eliminativism, then, if we have good reasons to reject phenomenal data, we should believe eliminativism.[15]

D. Conditional Eliminativism

The conclusion of the *Argument from Mysteries* is *Conditional Eliminativism*: If we have good reasons to reject phenomenal data, we should believe eliminativism.

Some realists reject *Conditional Eliminativism*. One way to do it is to deny *Explanation*. Some strong realists do so. They think that realism accounts for non-phenomenal data better than eliminativism; e.g., because something like the phenomenal powers view is correct. Some weak realists, like logical behaviorists and analytical functionalists, also deny *Explanation*. They think that phenomenal concepts are behavioral or functional concepts, and that eliminativists deny behavioral or functional properties needed to explain much physical data. Some other weak realists deny only *Mystery*.[16] They think that phenomenal consciousness does not really seem mysterious, or not in a way stemming from our concepts' characterizations.

Still, many realists concede *Conditional Eliminativism*. They simply insist that we have phenomenal data. For instance, debating Dennett, John Searle stressed that:

> The subjective feelings are the *data* that a theory of consciousness has to explain.[17]

[15] *Razor* resembles *theoretical conservatism* (the idea that, when possible, we must avoid innovations to deal with unaccounted phenomena), which Frankish (2016, 24) uses to argue for illusionism. Theoretical conservatism gives an edge to eliminativism only against strong realism, but *Razor* tilts the balance for eliminativism against both weak and strong realism: both face the apparent mysteries of consciousness.

[16] Analytic functionalists and logical behaviorists arguably deny both *Mystery* and *Explanation*.

[17] Searle (1997, 99).

Chalmers and Levine similarly rejected eliminativism only because they accepted phenomenal data:

> Eliminativism about consciousness is an unreasonable position *only* because of our own acquaintance with it. If it were not for this direct knowledge consciousness could go the way of the vital spirit [. . .]. Our knowledge that conscious experience exists derives primarily from our own case, with external evidence playing at best a secondary role.[18]

> Of course if [data about behavior] are your data, then conscious experiences themselves [. . .] become legitimate only as explanatory posits. Once we see how to account for the data without qualia, their legitimacy is undermined. I maintain, however, that conscious experiences themselves [. . .] are the primary data to which a theory must answer.[19]

This line of reasoning is old. James Cornman, an early opponent to eliminativism, who labeled the view, defended it in the 1960s. He criticized Quine's idea that "mental entities such as sensations" can be eliminated because

> like electrons, they are theoretical entities postulated as parts of scientific explanations, but, unlike electrons, they are unnecessary for such explanations.[20]

Against Quine, Cornman stressed that "sensations" are not postulated entities: they are things we are aware of, and

> What we are aware of is not postulated.[21]

This amounts to insisting that we have phenomenal data. American philosopher D.C. Williams already made a similar point in the 1930s when criticizing eliminativist interpretations of behaviorism.

[18] Chalmers (1996, 102).
[19] Levine (2001, 134). My *Argument from Mysteries* resembles what Levine (2001, 131) calls "accusations of theoretical irrelevance," one of the five eliminativist strategies he discusses. I agree that this strategy is insufficient.
[20] Cornman (1968a, 45).
[21] Cornman (1968a, 61).

> My consciousness, in sum, far from being "assumed" and "unobservable," as some behaviorists complain [. . .], is the only thing wholly "empirical" and immediately "observable" [. . .], and far from my wanting things "explained with consciousness left out," as some behaviorists offer, my conscious experience is what primarily I want explained.[22]

Conditional Eliminativism seems plausible to many realists, who simply deny its antecedent. The crux of the debate is: Do we have good reasons to reject phenomenal data?

3.3. Introspection and phenomenal data

Before turning to my argument against phenomenal data, some clarifications are needed.

First, I assume that the best candidate source for phenomenal data is introspection. If we have phenomenal data, we have introspective phenomenal data. If we do not have introspective phenomenal data, we do not have phenomenal data.

Second, we must clarify what the rejection of phenomenal data entails. It does not mean denying that, in introspection, it psychologically and epistemically seems to us that we have phenomenal experiences, and that, for this reason, prima facie, we have phenomenal data. Still, ultimately, we have good reasons to reject phenomenal data. This juxtaposition is unsurprising. We often start with data we later repudiate. A biologist can use a microscope and make observations she first legitimately treats as data, before correctly discarding this 'data' when realizing the microscope is defective. A detective justifiably treats a testimony as data until discovering that the mafia blackmailed the witness.

Rejecting phenomenal data also does not require claiming that phenomenal introspection is uninformative. Phenomenal introspection reliably tracks but misrepresents quasi-phenomenal states. Correctly interpreted, it provides *non-phenomenal data* about those. Still, rejecting phenomenal data goes beyond stating that phenomenal introspection *sometimes* errs: it requires rejecting *all* phenomenal data.

[22] Williams (1934, 465).

Third, we ask whether we have *good reasons* to reject phenomenal data, but we must clarify our dialectical starting point. Is the view that we have phenomenal data the default? How good must our reasons be to count as good reasons here? I just admitted that introspection makes it epistemically seem to us that we have phenomenal experiences: the view that we have phenomenal data has a head start. How considerable is it? Here are two reasons to treat it as modest.

One is that phenomenal introspection sometimes errs and generates false beliefs. For instance, based on introspection, we might believe that our visual phenomenal field is colored through and through, but this is wrong: further attentive introspection suggests that we do not phenomenally experience peripheral colors. Consider also the fraternity case.[23] A student is initiated into a fraternity. He is shown a razor, blindfolded, and told that the razor will cut his throat. He feels something on his neck, screams, and thinks he feels pain. One second later, he realizes he only felt cold, not pain: an icicle, not a razor, was on his throat. Arguably, the student wrongly believed, based on introspection, that he experienced pain. Besides, in various complex cases—e.g., regarding whether dreams are colored, or regarding whether our stream of experience is abundant and sparse—subjects make contradictory introspective reports, which probably cannot all be correct.[24] At least some subjects must be making introspective mistakes.

So, phenomenal introspection is *fallible*. Some go further and think it is *unreliable*,[25] but even if we simply think that it is fallible, it becomes hard to completely rule out that it could be *systematically* wrong about *some* things, like the existence of phenomenal experiences, even if it might be fairly reliable otherwise. If indeed we have non-negligible (even if very low) credence that phenomenal introspection is systematically mistaken in this way, the head start of the pro-phenomenal data view cannot be overwhelming. It cannot spare us the duty to examine the issue further.

There is a second reason to think that this head start is modest. Plausibly, introspection is an innate representational capacity, shaped by evolution through natural selection. However, *other* such capacities often systematically

[23] Pereboom (2011, 22–23). Pereboom borrowed this example from Christopher Hill (1991, 128–29), who himself took it from Rogers Albritton.

[24] Schwitzgebel (2011) argues for this at length. These examples are his.

[25] A venerable philosophical tradition (Comte [1830] 1989, first lesson; Cournot 1851, chap. 23), continued today by Schwitzgebel, insisted on introspection's unreliability.

mischaracterize features they reliably track. By induction, introspection is at least reasonably likely to do the same.

Consider perception. It presents us with qualities of perceived objects, including so-called secondary qualities, like colors. Plausibly, objects either do not really have secondary qualities, or these qualities are not as they appear. Vision presents us with colors, characterized as simple and primitive qualities of surfaces. Surfaces bear no such things, but only reflectance profiles grounded in their microphysical properties, which are neither simple nor primitive. Objects lack what Chalmers calls "Edenic colors," primitive and simple qualities, opposed to their "ordinary colors," which are not simple nor primitive.[26]

Various metaphysical options are available regarding colors. Maybe colors do not exist (color eliminativism). Maybe they do, but are not the simple and primitive properties they seem to be: rather, they are physically grounded reflectance profiles (color physicalism), or dispositions to produce representations of primitive colors (color dispositionalism).[27] Overall, what is plausible is that visual perception tracks properties that it mischaracterizes as simple and primitive qualities, and that color perceptions are reliable misrepresentations. The reasoning applies to secondary qualities represented in other modalities. This suggests that innate representational capacities shaped by evolution often mischaracterize what they reliably track.[28]

The same holds for domain-specific naïve cognition. Take naïve physics, a part of "core cognition" or "core knowledge":[29] a widely shared, arguably innate set of cognitive capacities that develops in early childhood and is partially conserved in adulthood. Naïve physics enables subjects to represent and predict their physical environment successfully. It also mischaracterizes the physical world: it represents that objects have (potentially curvilinear) "impetus," and thus sometimes take physically impossible trajectories.[30] It represents physical objects as spatially continuous, solid, bounded objects that do not spontaneously change,[31] which is false according to contemporary physics. Arguably, it also represents temporal relations, like simultaneity, as

[26] Chalmers (2006).
[27] For more, see Maund (2022).
[28] Pereboom (2011, 36–40) compared illusionism about primitive ("Edenic") colors and illusionism about the phenomenal. See also the next chapter.
[29] Carey and Spelke (1996); Spelke (2000); Spelke and Kinzler (2007).
[30] McCloskey (1983).
[31] Spelke (2000).

absolute, so that two spatially distinct events are either simultaneous or not, although simultaneity is not absolute according to Special Relativity. Still, naïve physics reliably tracks regularities approximately holding at our scales, which is probably why it was selected. It relies on reliable misrepresentations. Arguably, the same holds for naïve biology and other bits of core knowledge. This, again, suggests that innate representational capacities shaped by evolution often mischaracterize what they reliably track.

So, by induction, introspection is likely to mischaracterize what it otherwise reliably tracks. How strong is this inductive argument? This is debatable. In response, one could point out the disanalogies between introspection and perception or domain-specific naïve cognition. How such reference-class tennis ends depends on one's exact views of these capacities, but those receptive to the naturalistic perspective should not assume that introspection is radically different from other evolved representational capacities. They should give non-negligible credence to the idea that introspection systematically mischaracterizes what it tracks. The head start of the view that we have phenomenal data cannot be overwhelming.

Finally, one last thing must be clarified. The antecedent of *Conditional Eliminativism* is that we have good reasons to reject phenomenal data. How good must our reasons be to count as good reasons?

This depends on two things. First, our dialectical starting point, which I just discussed. If the pro-phenomenal data stance has a significant head start, our reasons must be stronger than if it does not. Second, our understanding of these 'good' reasons must be coordinated with the role these reasons play in the *Argument of Mysteries*. How good a reason must be depends on the weight given to the apparent mysteries of consciousness when assessing eliminativism's advantages. If eliminativism's avoidance of the hard problem and the explanatory gap, and more generally the fact that it does not have to deal with the apparent mysteries of consciousness, is a considerable theoretical gain, moderately convincing reasons to reject phenomenal data might be enough to make eliminativism preferable, and thus count as *good* reasons. If this theoretical advantage is small, decisive reasons to reject phenomenal data might be needed to prefer eliminativism: only such decisive reasons would count as *good* reasons. I argued, in Chapter 1, that the hard problem of consciousness and the explanatory gap are profound difficulties, plaguing our understanding of consciousness for centuries. Suppressing them yields significant theoretical gains. In my mind, strong but non-decisive reasons to reject phenomenal data count as good reasons here.

3.4. The Argument from Irrelevance

Here is the outline of the argument against phenomenal data.

Premise 1 ***(Local Explanation)*****:** Phenomenal realism does not account for phenomenal judgments better than eliminativism.

Premise 2 ***(Connection)*****:** If phenomenal realism does not account for phenomenal judgments better than eliminativism, phenomenal consciousness is not crucially relevant for phenomenal judgments.

Premise 3 ***(Relevance)*****:** If phenomenal consciousness is not crucially relevant for phenomenal judgments, it is probable that phenomenal judgments do not mirror phenomenal experiences.

Premise 4 ***(Mirroring)****:* If it is probable that phenomenal judgments do not mirror phenomenal experiences, we have good reasons to reject phenomenal data.

Conclusion ***(Against Phenomenal Data)****:* We have good reasons to reject phenomenal data.

The argument is valid. The terms it mobilizes were defined previously, except "phenomenal judgments," "mirrors," and "crucially relevant." More on this shortly.

This argument is a *debunking* argument. Debunking arguments typically rely on the idea that the epistemic status of our attitudes regarding certain entities is threatened if the entities do not play the right role in explaining these attitudes.[32] This one does not directly target the epistemic status of our attitudes,[33] but infers, from explanatory considerations, that certain correspondence between judgments about some entities and features of these entities is unlikely, which justifies that we discard putative data about these features.

Debunking arguments have been discussed, notably, about mathematics, modality, ordinary objects, God, and morality.[34] All such arguments are controversial. Debunking arguments are less controversial about extraterrestrial UFOs, the Yeti, or magic. In some controversial domains, many reject the conclusion of debunking arguments but grant the legitimacy of the

[32] Korman (2019).
[33] See Chalmers (2020a, n. 6) on non-epistemic debunking arguments.
[34] See Korman (2019, 2) for an exhaustive review.

debunking strategy. A pious Catholic priest might admit that *if* Jesus's resurrection does not explain his belief in the resurrection (e.g., because it does not explain the writing of the Gospel), his belief is debunked, but he would insist that it does explain his belief. Explicit debunking arguments about phenomenal consciousness have only been considered relatively late.[35]

I will first defend the premises of the *Argument from Irrelevance*, before explaining how it differs from neighboring arguments.

A. Premise 1—Local Explanation

Local Explanation states that phenomenal realism does not account for phenomenal judgments better than eliminativism.

Phenomenal judgments are the psychological processes by which we apply phenomenal concepts and judge that there are phenomenal experiences. *Introspective* phenomenal judgments are paradigmatic here, even if everything I say applies, *mutatis mutandis*, to non-introspective phenomenal judgments. As defined, phenomenal judgments are *physical entities*. They only have physical properties, captured by physical data.

Let us flesh this out. Suppose you have an intense headache. Based on introspection, you think, with a phenomenal concept, that you have a phenomenal experience of pain. This thought prompts you to do and say various things: e.g., it makes you say "eight" when a nurse asks you to rate your pain on a scale from one to ten. Whatever this thought's ultimate nature, it certainly has physical properties—it has properties pertaining to structures and dynamics. Maybe you think it has non-physical properties too, but focus on its physical properties, and stipulate that "phenomenal judgments" refers to the *physical* processes, endowed with physical properties, involved in thoughts of this sort. We could call them "physically conceived phenomenal judgments," but this would be cumbersome.

That there are phenomenal judgments in this sense is uncontroversial. Interactionist dualists think that they are caused or realized by mental, non-physical processes. Neutral monists view them as constituted by *aspects* of processes that are neither physical nor mental. Idealists think they are

[35] The first explicit illusionist debunking argument was presented (but not endorsed) by Chalmers (1996, 186–93; 2018, 44–49). Frankish (2016, 27; 2019) later endorsed a version of Chalmers's argument.

grounded in something mental and non-physical. Still, everyone agrees that they exist. About them, we have much non-tendentious (physical) data.

By definition, phenomenal judgments only have physical properties, captured by physical data. Therefore, *Local Explanation* follows from *Explanation*, the already-defended premise of my *Argument from Mysteries*, stating that phenomenal realism does not account for non-phenomenal data, including physical data, better than eliminativism. No new argument is needed for *Local Explanation*.

Like *Explanation*, *Local Explanation* is deniable. Strong realists for whom realism accounts for non-phenomenal data better than eliminativism (some interactionist dualists, and proponents of the phenomenal powers view) deny it. Logical behaviorists and analytical functionalists, who view phenomenal concepts as behavioral or functional concepts, also deny it.

Local Explanation is weaker than *Explanation*: objections to *Local Explanation* are objections to *Explanation*, but the reverse is not true. Still, some objections to both become salient only about *Local Explanation*. Since phenomenal judgments are tightly linked to phenomenal experiences, one could doubt that they can be explained as well if we do not appeal to phenomenal experiences as if we do. This threatens both *Local Explanation* and *Explanation*, but this becomes most salient when discussing the first. The solution to the illusion problem I later defend will hopefully show that we can explain phenomenal judgments as well if we do not appeal to phenomenal experiences as if we do.

B. Premise 2—Connection

Connection states that, if phenomenal realism does not account for phenomenal judgments better than eliminativism, phenomenal consciousness is not crucially relevant for phenomenal judgments. "Crucially relevant" means the following: A is crucially relevant for B when, if A had been absent or quite different, B would have been absent or quite different. A can be crucially relevant for B in virtue of causal or constitutive relations, be they direct or indirect.

The premise connects epistemological matters, regarding explanatory quality, to metaphysical matters, regarding dependence relations. It is a particular case of a plausible *Connecting Principle* stating that, if views asserting the existence of A do not account for B better than views denying it, A is not

crucially relevant to B. This seems plausible in ordinary and scientific cases. If A is crucially relevant to B, this should show up in our explanations: denying A should damage our explanations of B.

Take an ordinary case. The pulling of a gun's trigger was crucially relevant to the shot. Someone denying that the trigger was pulled will struggle to explain the shot. They will resort to ad hoc suppositions (e.g., the gun was dysfunctional and fired all by itself), typically yielding explanations inferior to those on which the trigger is pulled—e.g., diminishing the coherence of the explanation with other admitted facts about the gun's reliability and the soldier's actions.

Take a scientific case. Assume that the curvature of spacetime by the sun's mass plays a crucially relevant role in bringing about the 'anomalous' precession of the perihelion of Mercury. Views positing this curvature, like General Relativity, explain this precession better—more simply, more elegantly, and more coherently with our other well-justified views—than those that do not, like Newtonian mechanics.

Connecting Principle is modest, and weaker than what we could call *Strong Connecting Principle*: the principle that, if an entity is crucially relevant to a phenomenon, *mentioning it is needed* to explain the phenomenon. *Strong Connecting Principle* encounters plausible counterexamples that do not threaten *Connecting Principle*. Tables are crucially relevant to our beliefs about them, as they contribute to causing them, but we do not need to mention tables to explain these beliefs. We can explain them by mentioning only the properties of our brains and without mentioning our environment. We can also explain them by describing our environment, but without using the concept TABLE—e.g., by mentioning only atomic configurations.[36]

These are counterexamples to *Strong Connecting Principle*, not to *Connecting Principle*. We can explain our table-beliefs without mentioning our environment, but this explanation is inferior to those that mention our environment, notably when it comes to simplicity and elegance, as it requires many unexplained posits regarding sensory input. We can also explain our table-beliefs without mentioning tables, by describing only atomic configurations; fully developed, the explanation might be as good as it gets. However, conjoined with the *negation* of tables, the account becomes incoherent, given plausible implicit conceptual truths (e.g., if a thing is

[36] Chalmers (2018, 46).

table-shaped, has the right causal powers, and the right causal history, it is a table). So, tables are crucially relevant to our beliefs about tables, and views denying the existence of tables explain our table-beliefs less well than those admitting it. These examples are a problem for *Strong Connecting Principle*, not for *Connecting Principle*.

Connecting Principle is plausible. Therefore, so is *Connection*. Still, one could grant that *Connecting Principle* is usually correct, but not in the case of consciousness. If consciousness is an exception, the exception must be accounted for. One way to do so is to contend that phenomenal concepts are peculiar because, while their referent is crucially relevant to phenomenal judgments, denying that these concepts are satisfied nevertheless does not damage our explanations of phenomenal judgments. If phenomenal concepts are such, we can reject *Connection*. In some views, associated with type-B materialism,[37] and the so-called Phenomenal Concept Strategy,[38] phenomenal concepts are like this. They refer, *via* causal/informational or teleological factors, to physical states which are crucially relevant to phenomenal judgments. Still, they do not grasp their referent *via* any characterizing cognitive content, and explanations of phenomenal judgments can be indifferently conjoined with the affirmation or negation of the existence of phenomenal experiences. If we endorse a view like this, we naturally reject *Connection*.[39]

These views are popular, which is why *Connection* is probably the premise most weak realists would deny. However, I think that these views are plainly false. They fail to capture how our phenomenal concepts work. In Chapter 1, I stressed that phenomenal concepts provide a substantive characterization of their referents. Mystery intuitions follow from this characterization. Proponents of these views of phenomenal concepts, which enable one to reject *Connection*, state that phenomenal concepts do not substantively characterize their referents, and certainly not in a mysterious way that might threaten the capacity of physical states to satisfy these concepts. So, I rule out this sort of rejection of *Connection*. I return to it in Chapter 7. For now, let us accept *Connection*.

[37] Chalmers (2002a).

[38] Loar (1990; 1997); Papineau (2002); Tye (2003); Carruthers (2004); Balog (2012b); for the expression, see Stoljar (2005).

[39] These views were primarily developed to deny that zombies' conceivability entails their possibility. The issues are related: if zombies are inconceivable, eliminativism should struggle to explain phenomenal judgments without contradiction.

C. Premise 3—Relevance

Relevance states that, if phenomenal consciousness is not crucially relevant for phenomenal judgments, it is probable that phenomenal judgments do not mirror phenomenal experiences.

"Phenomenal judgments mirror phenomenal experiences" means that, in normal cases, there is a fine-grained partial correspondence between (aspects of) phenomenal judgments and (aspects) of phenomenal experiences. The mirroring relation does not require a one-to-one correspondence between these aspects. The correspondence can *sometimes* fail to occur, as long as it normally occurs.

This mirroring relation is extremely intuitive. Suppose you experience a sudden headache after lunch. In your diary, you write that the pain starts around 1 p.m., grows in intensity, comes with shoulder and neck pain starting around 3 p.m., and stops about half an hour after you take a painkiller at 4 p.m. Intuitively, what you write in your diary partially mirrors the quality, intensity, duration, and felt location of the felt pain. Writing patterns, which themselves depend on judgment patterns, mirror phenomenal patterns. Of course, the mirroring is partial: not every aspect of your experience is mirrored in your writing, and vice versa.

The mirroring relation is a matter of correlation: it is a subpart of what are traditionally called "psycho-physical correlations". It requires no concrete causal or constitutive relation between phenomenal experiences and what mirrors them: the relation can hold even if consciousness *is not crucially relevant in bringing about phenomenal judgments*. Still, since this correspondence is relatively fine-grained and concerns many instances, it is very implausible that it holds without phenomenal consciousness being crucially relevant to phenomenal judgments.[40] Let us see why with an example.

Suppose my sister calls and says: "Hi François, how are you?" In *Normal Phone Call*, everything goes as planned: there is a fine-grained correspondence between the words I hear and those she says. The first mirrors the second. The correspondence holds because of an appropriate *causal* relation between the processes (mediated by antennas, satellites, digital processing,

[40] If we admit arbitrarily contrived rules of correspondence, any sufficiently complex X mirrors any Y, which makes the mirroring notion useless. To make it useful, we must limit to ourselves to sufficiently natural rules, given the content of the entities involved (since, in the cases I examine, which involve utterances and judgments, the relevant entities are contentful). It is very hard to say positively what such rules are exactly, and I will not try here.

etc.): what my sister says is crucially relevant to what I hear. The causal relation makes the mirroring happen.

Consider other possible cases. In *Third Party Call*, our father told my sister to dial me at time *t* and say "Hi François, how are you?" My facetious father tampered with my phone to make it unreachable, while setting it up to ring and play the soundtrack of my sister saying these exact same words at time *t*. Here, a third party entered into some causal relation with two unrelated processes and ensured their correspondence even if the first is not crucially relevant to the second.

In *Coincidence Call*, my sister dials my number and says the words, but she does not reach me—the network is down. Still, out of sheer coincidence, my phone self-activates and makes random sounds auditorily indistinguishable from my sister saying: "Hi François, how are you?". This is possible, but comically unlikely.

In *Third Party Call* and *Coincidence Call*, a correspondence occurs even if what my sister says is irrelevant to what I hear. In both cases, the correspondence is due to luck; pure luck for *Coincidence Call*, luck that a third party played the right role in *Third Party Call*.

Come back to phenomenal judgments. In most views, phenomenal consciousness is crucially relevant to phenomenal judgments: it causes or constitutes phenomenal judgments, and this makes the correspondence happen. That this correspondence normally obtains out of pure coincidence is comically implausible: it is as if *all of our phone calls* were like *Coincidence Call*.

That this correspondence obtains because of a third party means that our situation is as if *all our phone calls* are like *Third Party Call*. It is hard to think of a third party capable of this, except a godly being or a complex set of laws of nature. However, that any of these two things ensures the correspondence is implausible. The existence of sufficiently fine-grained laws of nature ensuring this coincidence without direct relations between phenomenal experiences and judgments—which, e.g., epiphenomenal dualism probably has to posit—would be very lucky, although maybe not as much as pure coincidences. Appeals to God seem preferable: the many actions needed to ensure the correspondence could be explained in a unified manner by God's intentions. That God plays a role of this kind seemed plausible to some. Malebranche or Leibniz thought God himself ensured the correlation between minds and bodies, and therefore, between phenomenal judgments and phenomenal experiences. For Leibniz, God, during

Creation, pre-established harmony between minds and bodies, which at a deeper level are modifications of mental entities—monads.[41] Malebranche, like other post-Cartesians such as Cordemoy and La Forge, believed in "occasionalism," on which God is the only genuine cause of phenomena, and psycho-physical causal relations (my pain experience making me say "ouch") are merely apparent (*God* causes me to say "ouch," the pain experience being a mere "occasional" cause).[42] Some still find those ideas attractive. For Robert Adams,

> it is a theoretical advantage of theism that it makes possible a [theological] explanation" of the "correlation between phenomenal qualia and physical states.[43]

Still, I think that, if we do not operate with antecedent sympathy for the view that there is a God willing to ensure this correspondence, the view is implausible, which is why I rule it out.

To summarize, it seems very unlikely that phenomenal judgments mirror phenomenal experiences even if phenomenal consciousness is not crucially relevant for phenomenal judgments. Therefore, if phenomenal consciousness is not crucially relevant for phenomenal judgments, it is probable that phenomenal judgments do not mirror phenomenal experiences. This is *Relevance*.

D. Premise 4—Mirroring

Mirroring states that, if it is probable that phenomenal judgments do not mirror phenomenal experiences, we have good reasons to reject phenomenal data.

Why? Suppose you deny that the mirroring obtains. So, you deny that people's utterances mirror their phenomenal experiences.[44] My friend says

[41] Leibniz ([1714] 1989b).

[42] Occasionalism does not merely concern psycho-physical causation, but it was developed in discussions of psycho-physical causation (S. Lee 2020). Note that occasionalism does *not* entail that phenomenal consciousness is not crucially relevant for phenomenal judgments: it simply leaves this possibility open, although its standard form is arguably better interpreted as ruling it out.

[43] Adams (1987, 250). For another recent argument in this direction, appealing to the concept of psychophysical harmony, see Cutter and Crummett (2025).

[44] Utterances could mirror phenomenal states even if phenomenal judgments do not, but it would be easy to target this with a revised argument.

that she feels pain, but this does not correspond to how she phenomenally feels. I wrote yesterday in my diary that I was euphoric, but this does not correspond to what I felt. I am saying, right now, that I have a visual experience of white, but the words coming out of my mouth do not mirror what I experience. Can you believe all this, but maintain that you still have phenomenal data? Yes, but you need to accept a bizarre view of reality, on which you know about your phenomenal experiences—you legitimately take them to exist with such and such features—but do not normally truthfully say, write, or (physically) think about them.

If this view is correct, a chaotic skeptical scenario is actual. The scenario is skeptical, as it depicts a reality that massively departs from what we think we know, and we have no positive reason to believe it. It is chaotic, as it depicts a world in which there are entities we are familiar with—phenomenal experiences—but the basic regularities we take for granted about them do not hold.[45]

Contrast such a chaotic skeptical scenario with a familiar skeptical scenario. Suppose phenomenal solipsism is true: nothing exists but your own phenomenal experiences. Still, regularities between patterns of phenomenal experiences might hold. You feel pain—Phenomenal Experience 1 (PE1). Your PE1-like states are regularly followed by phenomenal experiences of an intention to say that you are in pain—PE2-like states. PE2-like states are then regularly followed (say) by auditory experiences of yourself saying "I am in pain"—PE3-like states, etc. This regular chain of complex phenomenal experiences *grounds* certain structural and/or causal features, in virtue of which you make phenomenal judgments. These phenomenal judgments, in turn, partially mirror your experiences: aspects of the (structural and/or causal) pattern instantiated by PE2 states and PE3 states mirror aspects of PE1 states. This scenario is skeptical, but it is not chaotic. In this scenario, you have phenomenal judgments, phenomenal experiences, data about phenomenal experiences, and phenomenal judgments mirror phenomenal experiences.[46]

Now, in contrast, to think of a world where you have phenomenal data even if phenomenal judgments do not mirror phenomenal experiences,

[45] By contrast, illusionism is not a skeptical scenario. It offers a picture of reality departing from what we usually take for granted, but there are reasons to accept it. It is not chaotic either, since it does not posit entities about which the regularities we take for granted do not hold.

[46] Depending on whether the right causal or constitutive relations between experiences occur, experiences might or might not be crucially relevant for phenomenal judgments about them.

you must imagine something more radical, like a chaotic solipsistic world. In this world, only your phenomenal experiences exist, without appropriate regularities between them. You have phenomenal experiences, and you somehow have data about them. Still, your experience of hearing yourself say "I am in pain" does not usually correspond to an earlier experience of pain. You have phenomenal data, but phenomenal judgments do not mirror phenomenal experiences.

We are not in a world like this. If we have phenomenal data, our phenomenal judgments mirror our experiences. Therefore, if the mirroring relation probably does not hold, we probably do not have phenomenal data. If the possession of phenomenal data is improbable, we have *good reasons* to reject phenomenal data.[47] We reach *Mirroring*: if it is probable that phenomenal judgments do not mirror phenomenal experiences, we have good reasons to reject phenomenal data.

E. Against Phenomenal Data

The argument's conclusion is that we have good reasons to reject phenomenal data.

We make phenomenal judgments. It seems to us, psychologically and epistemically, that we have phenomenal experiences and possess phenomenal data, but ultima facie, we have good reasons to discard our putative phenomenal data. To plausibly maintain that we have phenomenal data, we must claim that our phenomenal judgments mirror phenomenal experiences (*Mirroring*). This is improbable if phenomenal consciousness is not crucially relevant to bring about phenomenal judgments (*Relevance*). However, phenomenal consciousness is not crucially relevant to bring about phenomenal judgment, since phenomenal realism does not account for phenomenal judgments better than eliminativism (*Connection* and *Local Explanation*).

Each premise of the argument is deniable. Weak realists attracted to something like the Phenomenal Concept Strategy reject *Connection*, which is false in some views of phenomenal concepts. Logical behaviorists and analytical functionalists reject *Local Explanation*. So do dualist interactionists and proponents of the phenomenal powers view. Some occasionalists, believers in a pre-established harmony, and epiphenomenal dualists, might

[47] But see previously discussed quibbles about which reasons are good reasons here.

reject *Relevance*. Finally, disagreement can arise about what counts as a good reason to reject phenomenal data. One could argue that only overwhelming reasons to reject phenomenal data are *good reasons* to do so, and thus deny *Mirroring*. Still, all premises of the argument are prima facie fairly plausible.

3.5. Reaching the eliminative claim

The *Argument from Mysteries* ended in a stalemate. Many eliminativists and realists agreed that, phenomenal data set aside, eliminativism is preferable to realism. They disagreed on whether we really have phenomenal data.

The *Argument from Irrelevance* shows a way out of the stalemate. It exploits the fact that, to maintain that we have phenomenal data, we have to recognize certain relations between some physical processes—phenomenal judgments—and phenomenal experiences. We can then consider physical data about phenomenal judgments, and notice that this data does not look like what it should look like if the right relations between phenomenal judgments and phenomenal experiences occurred. This provides a good reason to reject phenomenal data. Since we have good reasons to reject phenomenal data, and given the conclusion of the *Argument from Mysteries*, we should believe eliminativism.

The *Argument from Mysteries* and the *Argument from Irrelevance* work in tandem. Both are needed to support the eliminative claim. On the one hand, the fact that phenomenal consciousness seems mysterious, and fails to help us explain non-phenomenal data, puts pressure on us to deny its existence. However, this pressure would have to be resisted if we had convincing direct data for its existence. The *Argument from Mysteries* alone does not get us to eliminativism.

The *Argument from Irrelevance* alone does not do it either. Good reasons to discard phenomenal data would not be enough to deny phenomenal consciousness if it were a useful explanatory posit and helped us explain the non-phenomenal. Even granting phenomenal consciousness is not explanatorily useful, if it did not *seem mysterious*, and if it were smoothly integrated into our best picture of reality, we would need *decisive* reasons to discard phenomenal data in order to properly support the eliminative claim. It is doubtful that the *Argument from Irrelevance* gives such decisive reasons: its premises are plausible, but deniable. However, precisely because phenomenal consciousness

seems mysterious and does not integrate well within our current best picture of reality, there is pressure to deny its existence. This lowers the bar of what counts as *good reasons* to discard phenomenal data, and the *Argument from Irrelevance* can tip the balance.

So, assessing the eliminative claim, we must consider these two arguments together. Still, even this tandem is not decisive. It is only the first building block of the case for eliminativism.

3.6. Other debunking arguments

If you accept the *Argument from Irrelevance*, do you have to accept debunking arguments in *other domains*? Not necessarily. The debunking strategy taken by the *Argument from Irrelevance* appeals to *explanatory indifference*. It is not suited to debunk beliefs in ordinary objects, since denying the existence of, e.g., tables, damages our explanations of table-beliefs. Could it debunk beliefs about modality and mathematics? This depends on whether there are explanatory advantages of positing modalities and mathematical objects: if so, as is plausible, the strategy fails.

The explanatory indifference strategy seems more effective to debunk moral beliefs. Still, accepting the *Argument from Irrelevance* does not commit you to discarding moral data about, say, what is right and wrong. It certainly does not commit you to moral nihilism, on which nothing is right or wrong. For instance, if you think that moral truths (but not phenomenal truths) are a priori, you might claim that, while eliminativism is not explanatorily inferior to phenomenal realism, moral nihilism is explanatorily inferior to moral realism, as it features (possibly implicit) incoherences. The difference between the moral case and the consciousness case is plausible if we appreciate that zombie worlds seem positively conceivable, while moral nihilist worlds, in which torturing innocents just for fun is not wrong, do not.[48] Even if considerations of explanatory indifference give *some* reasons to reject both phenomenal and moral data (e.g., if you think moral truths are a posteriori), phenomenal eliminativism could be ultimately supported, but not moral nihilism. Phenomenal consciousness might seem more mysterious than morality—e.g., if moral properties are easier to reduce metaphysically to the physical than phenomenal

[48] Chalmers (2020a, 266).

properties. If so, stronger reasons are needed to discard moral data than to discard phenomenal data.

How does the *Argument from Irrelevance* relate to other debunking arguments about consciousness? Sydney Shoemaker may have given the first such argument.[49] Assuming phenomenal realism and something like a causal theory of knowledge, he deployed a debunking strategy against anti-functionalism about phenomenal experiences, roughly claiming that, if phenomenal properties were non-functional, we would have no knowledge of them.

Debunking considerations were later key to Frank Jackson's repudiation of the Knowledge Argument. Jackson claimed, against his old self, that the nature of sensory processes is deducible from information about the physical nature of the world despite "the manifest appearance to the contrary that the knowledge argument trades on." He did so by appealing to the following principle:

> [Do] not have opinions that outrun what is required by the best theory of these opinions' causal origins.

He conjoined it with the following claim:

> [Our] knowledge of what it is like to see red and feel pain has purely physical causes.[50]

Contrary to Shoemaker's and Jackson's arguments, my *Argument from Irrelevance* appeals to explanatory considerations, not causal ones. It does not target our having knowledge of consciousness, but the correspondence between phenomenal judgments and phenomenal experiences required by our possession of phenomenal data. It ultimately supports a view about the inexistence of the phenomenal, not about its nature.

Chalmers was the first to present two explicitly illusionist debunking arguments, even if he rejected their conclusions.[51] His first argument concludes that our *beliefs* about phenomenal consciousness are not justified.

[49] Shoemaker (1975). See Chalmers (2018, n. 32) for the historical claim. Chalmers mentions other arguments with a debunking flavor against epiphenomenal dualism (Latham 2000; Yudkowsky 2015), or parallelism (Shaffer 1968).

[50] Jackson (1998b, 76–77).

[51] Chalmers (1996, 186–93; 2018, 44–49).

Its premises are (a) that we can explain these beliefs independently of phenomenal consciousness, and (b) that the existence of such an explanation undermines the justification of these beliefs. Chalmers pointed out the limits of the argument: consciousness could play an essential role in justifying our beliefs about it, saving their justified status, even if these beliefs can be explained independently of consciousness.

His second argument (the "coincidence argument") avoids this difficulty by focusing on physically understood *intuitions* about consciousness, and dropping reference to justification. It appeals to the premises that (a) if our intuitions regarding phenomenal consciousness are explained independently of consciousness, then their correctness is a coincidence; that (b) our intuitions are indeed so explained, and that (c) the correctness of our intuitions is not a coincidence. The argument concludes that these intuitions are incorrect.

My *Argument from Irrelevance* resembles the coincidence argument, with its focus on the explanation of physically understood phenomenal judgments. Still, I find it preferable to Chalmers's coincidence argument, for three reasons.

First, Chalmers's coincidence argument uses non-reductionism about consciousness as a background assumption, and does not target reductionist realism.[52] The *Argument from Irrelevance* has a wider scope. It only establishes illusionism in tandem with the *Argument from Mysteries*, but operates without non-reductionist assumptions.

Second, the *Argument from Irrelevance*, but not Chalmers's argument, addresses certain possibilities, in which we have phenomenal data even if our phenomenal judgments do not mirror phenomenal experiences. In these scenarios, illusionism is false and can be legitimately ruled out, but phenomenal judgments are mostly incorrect.

Third, and most importantly, Chalmers's coincidence argument mobilizes the claim that we can explain our intuitions independently of phenomenal consciousness. As Chalmers concedes, this is ambiguous between a claim of *descriptive* independence—we can explain intuitions without *mentioning* phenomenal consciousness—and *causal/constitutive* independence—we can explain these intuitions by appealing to elements causally/constitutively independent from consciousness.[53] Descriptive independence is plausible,

[52] Chalmers (2018, 45).
[53] Chalmers (2018, 46).

but does not grant the debunking conclusion, since the reasoning would face counterexamples: we can explain our intuitions about tables without mentioning tables, but these intuitions are not causally independent of tables, and we do not think that descriptive independence debunks our intuitions about tables. Causal/constitutive independence warrants the debunking conclusion, but it is not antecedently so plausible.

The *Argument from Irrelevance* escapes this dilemma by relying on *Local Explanation*. *Local Explanation* is plausible and weaker than a claim of causal/constitutive independence. Still, it goes beyond descriptive independence. It is about *explanatory indifference*, which is somewhere between descriptive and causal/constitutive *independence*. Its use for debunking does not lead to similar counterexamples. While phenomenal realism does not fare better than eliminativism in explaining phenomenal judgments, denying tables does damage our explanations of our intuitions about tables. *Connection* then links explanatory indifference to matters of causal/constitutive independence, making the relevant commitment explicit and plausible—though deniable.

At a deeper historical level, all these debunking considerations exploit the same apparent facts that preoccupied Early Modern philosophers discussing the mind–body relation in a post-Cartesian context: the mental exists, the physical exists, they are distinct, and yet systematically correlated. How can this be?

The main options in the metaphysics of mind are ways to make sense of these appearances harmoniously. Illusionism grants that there is a sense of "mental" such that the mental *is* distinct from the physical. However, in this sense, the mental does not exist, but only seems to exist. The *Argument from Irrelevance* points out that a subset of the physical—phenomenal judgments—would have to have certain relations with these apparent peculiar mental aspects if we were to have data about them, and stresses that our explanations of this subset do not look the way they should look if indeed these relations obtained, giving us good reasons to discard putative data about these mental aspects, and ultimately allowing us to reject their existence.

* * *

The *Argument from Mysteries* shows that, if we have good reasons to reject phenomenal data, we should believe eliminativism. The *Argument from Irrelevance* shows that we have good reasons to reject phenomenal

data. The final conclusion is that we should believe eliminativism. Of course, ultimately, the assessment of the eliminative claim must be holistic. It requires discussing objections (see Chapter 7), normative issues (see Chapters 9–10), and the illusion problem (Chapters 4–6), to which I now turn.

4

Approaching the House of Mirrors

The Illusion Problem

4.1. Understanding the illusion problem

Why does phenomenal consciousness appear to exist? Answering this question means solving the *illusion problem*. It is the second main challenge faced by illusionists. As Keith Frankish puts it:

> [T]he task [. . .] is to explain our illusory representations of phenomenality, not phenomenality itself, and the hard problem is replaced by the illusion problem.[1]

In this chapter, I present the problem and criticize existing candidate solutions. In the next chapter, I present my solution.

A. Why solve the illusion problem?

Why explain the illusion of phenomenal consciousness? First, because illusionists are committed to the illusory appearance of phenomenal consciousness. They must explain the existence of their posit. In this respect, they are not unlike realists, who must solve their own distinct *knowledge problem*. The *appearance problem*—explaining why phenomenal consciousness seems to exist—is the neutral one that everyone must solve.

Second, illusionists have *dialectical* reasons to solve the illusion problem. For instance, illusionism purports to be the best overall account of our data, but this can only be appreciated once we see that it can explain the appearance of phenomenal consciousness. Besides, solving the illusion problem plausibly requires explaining phenomenal judgments in an eliminativist

[1] Frankish (2016, 11).

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0005

framework, which strengthens the case for *Local Explanation*, a premise of the illusionist argument. Moreover, assuming that the illusion problem is solvable, different candidate solutions still plausibly have different implications regarding how prima facie justified phenomenal realism is, and how epistemically blamable realists are. Our exact solution to the illusion problem is dialectically relevant: our illusionism is more convincing if our view makes the realist beliefs false but epistemically excusable.

Third, solving the problem eases a concern. Realists often think that illusionism is incoherent, because the appearance of phenomenality requires phenomenality. As noted in Chapter 2, illusionists only claim that phenomenal consciousness appears to us in a psychological sense, and maybe in an epistemic sense, but not in a phenomenal sense. Their view is coherent. Still, one could fear that even the psychological appearance cannot be coherently explained by illusionists. Solving the illusion problem addresses this concern.

B. The illusion problem's explanandum

Phenomenal consciousness seems to exist. We have shared introspective compelling dispositions to believe we have phenomenal experiences, represented with phenomenal concepts. This is the explanandum of the illusion problem.

To better understand it, we must determine what *phenomenal concepts* are. We do not yet need a view of their nature, but *criteria* to identify them, so that we can establish whether a view explaining a disposition to believe in Xs explains a disposition to believe in phenomenal experiences. To find such criteria, return to the definitional strategy deployed in Chapter 1.

Earlier, I mentioned examples of mental states: seeing a white page, smelling coffee, recalling a taste, and imagining a visual scene. I claimed that we latch onto phenomenal concepts by focusing introspectively on states like those. This gives us a criterion: phenomenal concepts are notably applied when introspecting *sensory perception*, or states derived from sensory perception like imagination and imagistic memory—in contrast, maybe, to other states, like beliefs.

I also distinguished phenomenal concepts from other types of concepts, such as concepts of access-consciousness and self-consciousness, physical and functional concepts, etc. I stressed that phenomenal concepts

characterize their referents as mysterious—obvious, irreducible, and undetectable—and that they allow us to distinguish, in their referents, a *qualitative* dimension (apparently ineffable and related, in sensory perception, to perceived sensible qualities) and a *subjective* dimension. All this provides corresponding criteria.

Therefore, solving the illusion problem means explaining a shared introspective compelling disposition to believe falsely that we enter states represented with certain concepts. These concepts must notably be applied when introspecting sensory perception and derived processes. They must be distinguishable from concepts of access-consciousness and self-consciousness, as well as, more generally, physical and functional concepts. They must characterize their referents as appropriately mysterious and allow for a distinction between qualitative and subjective character.

So, solving the illusion problem requires explaining why we tend to believe that we enter certain states intuited as mysterious: it implies solving Chalmers's *meta-problem of consciousness*. Finally, ideally, we should explain our intuitions of epistemic and ethical significance.

Using these criteria does not mean that we *define* phenomenal consciousness as an obvious, irreducible, and undetectable entity, distinct from access-consciousness and self-consciousness. What instead happens is this. Using various props, we latched onto our already-possessed phenomenal concepts. We defined phenomenal consciousness as what satisfies these concepts. Reflecting on the props we used, we inferred features of our phenomenal concepts. This gives us criteria to determine whether a view really explains a disposition to believe in phenomenal experiences.

This characterization of the illusion problem's explanandum echoes the claims of those who think that phenomenality is *truly mysterious*, like strong realists. This might seem suspicious. Strong realists are wrong: Should we really center their claims when delimiting our explanandum?[2] However, phenomenal consciousness appearing mysterious does not mean that we have a genuinely mysterious form of consciousness. Besides, we should expect the characterization of phenomenal consciousness provided by strong realists to reflect how phenomenal consciousness introspectively appears better than the one provided by weak realists. Weak realists think consciousness is not as it seems. Hence, they might not use introspection much to establish the nature of phenomenality, notably when its verdicts support strong realism,

[2] Rosenthal (2019a; 2022).

and their claims are less likely to match introspection's verdicts than strong realists'.

An analogy helps. Suppose we have records from two traditional societies: *Dreamers*, who thought dreams put us in contact with a higher reality, and *Practicals*, for whom they are meaningless brain disturbances. *Practicals* were maybe closer to the truth, but *Dreamers* were more likely to explore dreams and collect dream reports. Contemporary dream researchers would be more interested in what *Dreamers* said about dreams. The same goes for phenomenality. The claims of strong realists, who think consciousness exists as it introspectively seems, will better mirror phenomenal consciousness's introspective appearance, and our characterization of the *appearance* of consciousness will echo their claims about consciousness.

C. The explanans of the illusion problem

We search for an explanation of the illusion of phenomenal consciousness. What should our explanans look like? First, its existence must be compatible with illusionism. Second, describing our explanans should allow us to deduce features of our explanandum. Ideally, we should deduce its detailed, quantified features, but for now, a rough, informal deduction of non-quantified features will do.

Third, explanations are rarely *ultimate* explanations: an explanans often calls itself for an explanation and becomes an explanandum. This allows us to specify our desired explanans further. Our first explanandum is an introspective disposition to believe. To explain it, we need a *synchronic and psychological* explanation, referring to the processes responsible for the disposition: introspective processes. This is our first explanans. We can describe these processes at various levels—neural, algorithmic, representational, behavioral, etc.—as long as we do not mention phenomenal experiences.

Ideally, we must then explain *why* we have these processes: we should give a *diachronic and historical* explanation of our having them. These psychological processes then become our second explanandum; cultural, social, or evolutionary causes are our second explanans. As our first explanandum is a shared disposition, our first explanans is likely to be a shared psychological process. Its shared nature makes it naturally explainable by phylogenetic history and evolutionary considerations. Evolutionary explanations take various forms. Some explain processes as adaptations; others, as resulting from

random evolutionary events (like genetic drift, accounting for the evolution of blood groups). Yet others explain processes as non-adaptive by-products of adaptations—exaptations. Again, this diachronic and historical explanation must be compatible with illusionism.

4.2. The *Common Core*, AST, and QIH

I will not review all proposed explanations of the appearance of phenomenal consciousness.[3] I will focus on a *Common Core* shared by the candidate explanations I find the most promising, which will form the basis of my view. Then, I will show why these promising explanations fail to solve the illusion problem—why we need to go beyond the *Common Core*.

In my mind, the most promising explanations of the appearance of phenomenal consciousness were developed by Derk Pereboom and by Michael Graziano.[4] Their views share the following core.

> *Common Core*: Our introspection has hard-wired and innate features, making it represent some of our mental states as having special properties. However, these representations are inaccurate. We do not really enter states with these properties. These inaccurate introspective representations are phenomenal concepts. These special properties are *phenomenal properties*. Therefore, our innate, hard-wired introspection generates the fallacious appearance of phenomenal experiences, which explains the illusion of phenomenal consciousness. The illusion is widely shared and compelling. It is robust and cognitively impenetrable, somewhat like perceptual illusions.

Pereboom's view is the "qualitative inaccuracy hypothesis" (QIH), on which introspection systematically misrepresents our real phenomenal states as having phenomenal properties gifted with a qualitative nature they lack. As introspection is hard-wired, the misrepresentation is robust and persists, irrespective of what we believe. (Pereboom does not say that these hard-wired features are innate, but it fits his view.) This version of QIH, which is Pereboom's, is a weak illusionist one, but a strong illusionist version is

[3] For overviews, see Frankish (2016, 16–20); Chalmers (2018, 19–35).

[4] Pereboom (2011); Graziano (2013). Pereboom does not endorse the view but sees it as an open possibility. To simplify, I speak of it as if it is his view.

possible, in which phenomenal properties *are* the inexistent qualitative natures. Strong illusionist QIH is the view I used to abstract the *Common Core*. Pereboom himself views the difference between the two versions as "verbal and non-substantive."[5]

Graziano's view is the "Attention Schema Theory" (AST).[6] For AST, our brain, in virtue of some innate trait,[7] monitors its own attentional processes via a schematic representation: the "attention schema." This self-monitoring recruits representational resources mobilized in social cognition when representing others' attention. The representation is a schema: instead of representing the complexities of attentional processes, it depicts a *simple* relation of "awareness" between a subject and an attended piece of information. The schema's features are robust: whatever we believe, it seems that we enter simple relations of awareness, grasped as "a fluidic substance, [. . .] an experience, [. . .] a sentience."[8] Arguably, these represented relations of awareness are no other than phenomenal experiences. However, we never enter these relations. There are no phenomenal experiences. As Graziano puts it:

> Consciousness doesn't happen. It's a mistaken construct.[9]

Graziano's and Pereboom's approaches differ. Graziano's is neuroscientific, whereas Pereboom's is philosophical. Pereboom focuses on *introspection*, which requires attention, effort, and mobilizes quasi-perceptual nonconceptual representations related to phenomenal concepts. Graziano focuses on automatic self-monitoring, an effortless process that consists in manipulating what might be read as conceptual representations.[10] I set these differences aside in the *Common Core*, which is my starting point.

Other illusionist views accept the *Common Core*.[11] Accepting it is natural if one views the appearance of phenomenal consciousness as a shared and compelling illusion. New-wave illusionists like Graziano and Pereboom, with their perceptualist take on the illusion of phenomenality, naturally embrace

[5] Pereboom (2011, 188).
[6] Graziano (2013).
[7] Graziano (2013, 208).
[8] Graziano (2013, 80).
[9] Graziano (2016). He later opted for verbal weak illusionism (Graziano 2019a). More on this in Chapter 7.
[10] This is how I read the "descriptive information" Graziano (2013, 34, 70) ascribes to our monitoring process.
[11] Humphrey (2011).

it. Theoretical eliminativists, with their mistakist stance, viewing the illusion of phenomenal consciousness as resulting from flawed reasoning, typically reject it. For instance, Dennett thought that the illusion of phenomenal consciousness is not mainly explained by hard-wired features of introspection, but rather by a variety of factors, including cognitive mistakes, cultural traits, and early education.[12]

I think that the illusion of phenomenal consciousness is compelling. I am relatively close to new-wave illusionists and their perceptualist stance, and I find the *Common Core* a good starting point. However, as I will show, the *Common Core* is insufficient: QIH and AST do not solve the illusion problem, as they fail to explain the unique compellingness of the illusion of phenomenal consciousness. We have to move beyond the *Common Core*. While the illusion of phenomenal consciousness shares certain features with perceptual illusions, making them similarly robust, it is also uniquely strong, in a way that cries out for explanation and requires an exceptionalist approach.

4.3. The intuition of irreducibility

Solving the illusion problem requires explaining mystery intuitions. Can AST or QIH do it? Consider first the intuition of irreducibility: our shared, compelling, and conceptual intuition that phenomenal consciousness is epistemically and metaphysically irreducible.

A. The Attention Schema Theory and the problem with sparse-illusion views

Graziano's AST views the attention schema as representationally *sui generis*, which might explain the conceptual independence of phenomenal and physico-functional concepts. Graziano also hints at a possible explanation of significance intuitions and attempts to explain the intuition of undetectability.[13] The attention schema mechanism is evolutionarily explainable as an adaptation, since possessing an attention schema advantages a creature

[12] Frankish (2023) remains neutral: he thinks that the illusion of phenomenal consciousness might derive from introspection, but insists that introspection could also merely generate a "proto-illusion," leading to the illusion of phenomenality only when combined with certain beliefs.

[13] Graziano (2013, 99–103).

by helping predict, guide, and regulate attention.[14] Still, I will show that AST does not explain our intuition of irreducibility.

AST views our attentional schema as *incomplete*. For AST, the schema grasps complex attentional states as consisting of simple relations of awareness between a subject and a content. However, it is not *positively incorrect*. It fails to capture the nature and internal complexities of attentional processes, but it does not positively depict the instantiation of some non-existent extra feature, nor the non-instantiation of an existent feature. It *simplifies*. It lies by omission, not by commission.

Suppose that introspection relies on such incomplete representations. Why can we not simply admit that they are incomplete and possibly completed by science? The incompleteness of our attention schema does not explain why we resist completing the fragmentary picture it provides with what science says—why we resist admitting that what the schema represents is physical and explainable. To explain the intuition of irreducibility, it is not enough to suppose that our attention schema is incomplete. We must also suppose that we *believe* it to be complete: we believe that our internal states only have the properties depicted by this schema.

Graziano thinks that this is what happens. He compares consciousness and white light. Our perceptual system presents white light schematically and incompletely: it does not show that it is composed of all the colored lights of the spectrum. Historically, we thought our grasp of white light was *complete*—we thought it was really *pure* and *not composed of anything*. This is why people first rejected Newton's view that white light is a mixture of all other colored lights.[15] For Graziano, the same happens with consciousness. We view it as irreducible because our attention schema represents our internal states incompletely *and* we mistakenly believe the schema is complete.

The problem is that this explanation fails to explain the robustness of our intuition of irreducibility.[16] Suppose our irreducibility intuition depends on our belief in introspection's completeness. Without this belief, phenomenal experiences should cease to seem irreducible. Their explainable

[14] Graziano (2013, 208–14). Graziano thinks these schematic representations were later recruited to serve other adaptive functions: strengthening social cohesion or, following Humphrey (2011), "making life emotionally lovely and therefore worth living."

[15] Graziano (2013, 49, 80).

[16] For a first version of this criticism, see Kammerer (2022b).

and purely physical nature should become intuitive. Similarly, when we dropped the belief that our grasp of white light is complete, Newton's view became intuitively plausible. Therefore, this explanation implies that our intuition of irreducibility is cognitively penetrable—vulnerable to changes in beliefs about introspection's completeness. It implies that it is not robust, and probably not widely shared, since many of us think introspection is incomplete. However, the problem is that this intuition *is* shared and robust; at least, this is what I presuppose for now. In this respect, AST's explanation of the irreducibility intuition faces the same problem as the explanation of theoretical eliminativists, which cannot account for the robustness of this intuition.

One way to fix the problem is to posit a shared and compelling intuition that introspection is complete. Together with the incompleteness of introspection, this would explain a shared and compelling intuition of irreducibility. Neither Graziano nor those defending similar explanations take this route, which incidentally comes close, in some respects, to the view I later defend.

Graziano, facing an early version of this objection, conceded that it is "insightful."[17] In response, he added that the attention schema is not only incomplete and silent on the phenomenal's physicality. It also represents phenomenal experiences as states for which *physical properties* are *irrelevant*—while falling short of representing them as non-physical.

This answer is insufficient. As Chalmers remarked,[18] we understand various processes, like computational processes, as such that their physicality is irrelevant. The physical realizer of a computation does not matter to the computation as such. Still, the physical realization of computational processes is not counterintuitive.

To explain the irreducibility intuition, we must go beyond AST. More generally, we must go beyond "sparse-illusion views,"[19] in which the incomplete nature of introspection explains our irreducibility intuition. We must grant that introspection *positively* characterizes phenomenal experiences as having (or lacking) certain features whose presence (or absence) would make phenomenal experiences irreducible. We should accept a "rich-illusion view."

[17] Graziano (2019a, chap. 7).
[18] Chalmers (2020b, 212–13).
[19] See Kammerer (2022b) for the first use of this terminology.

B. QIH, PIH, and rich-illusion views

Pereboom's QIH is a rich-illusion view. For QIH, introspection represents phenomenal properties gifted with a qualitative nature they in fact lack: it involves a *positive* misrepresentation. If we had phenomenal properties with this qualitative nature, they would be epistemically and metaphysically irreducible. So, whether or not we believe introspection to be complete, we are disposed to believe we enter states with these qualitative phenomenal properties: we have a robust tendency to believe in states that are irreducible, epistemically and metaphysically.

What is this *represented* but *non-existent* qualitative nature? Pereboom suggests it could be a primitive nature. For Pereboom, a property is represented as *primitive* when it is represented (a) as having a qualitative nature entirely presented in our representation and (b) as having a nature that is metaphysically simple, i.e., not constituted by, or analyzable into, multiple properties.[20] The primitivist inaccuracy hypothesis (PIH) is a version of QIH stating that introspection represents phenomenal properties as primitive.

Why would introspection represent phenomenal properties as primitive? There are plausible explanations. A system representing highly complex properties with no information about their underlying nature, and with no need to represent this nature, could do so by 'positing', so to speak, a set of qualities it ascribes to objects in its representation. The underlying nature of the properties being irrelevant, the system could conveniently represent these qualities as primitive, not composed of (or analyzable into) anything, and devoid of any extra hidden nature.[21] About the highly complex properties it tracks and represents as primitive, the system could make difference judgments ("X is different from Y"), identity judgments ("X is the same as Y"), and differential resemblance judgments ("X resembles Y more than it resembles W but less than it resembles Z"). It would then implicitly place these qualities within a 'quality space' matching these discriminations.[22]

Visual perception might precisely do this with colors and represent them as "Edenic colors,"[23] even if in reality there are no Edenic primitive colors,

[20] Pereboom (2011, 16–18).
[21] Chalmers (2018, 26–27). See also Chalmers (2006); Hall (2007).
[22] Approaching sensible qualities in terms of quality space is a popular option (Rosenthal 1999; 2010; for a summary and an attempt at using this approach to explain irreducibility and ineffability intuitions, see Weisberg 2023, 124–29).
[23] Chalmers (2006).

but only non-primitive ordinary colors: highly complex physical properties finely detected, discriminated, and classified by the system. Pereboom's suggestion is that introspection does the same with phenomenal properties. This goes one step beyond what Graziano's AST says.

Plausibly, PIH explains an intuition of irreducibility. Representing phenomenal properties as primitive, we represent them as having a *simple* essence captured in our representations. As phenomenal concepts are independent of physical concepts, our representations of phenomenal properties do not characterize them as having a physical essence. Therefore, we represent them as irreducible to the physical. More generally, we represent them as irreducible to anything substantively conceived in a way that is not clearly phenomenal. PIH also explains that we think of phenomenal properties as epistemically irreducible. Arguably, the emergence of primitive properties in the mental domain challenges scientific explanations, which typically deduce complex composed phenomena from more basic physical phenomena.

Moreover, PIH might account for our intuition of ineffability. If our representations of phenomenal properties place primitive phenomenal properties in a quality space corresponding to the discriminations we make, these representations have rich implicit informational content. We implicitly represent phenomenal properties' positions in the quality space, something we cannot explicitly access and verbally express, but only indirectly and partially elicit by making specific judgments. Much of what we know about these properties will seem ineffable.

Finally, PIH views these intuitions as resulting from the way our phenomenal concepts, deployed via a hard-wired introspective module, characterize phenomenal properties. It explains that they are shared, compelling, and conceptual.

PIH handles the irreducibility intuition better than AST. This is no surprise: rich-illusion views, viewing introspection as involving positive misrepresentations, are better positioned than sparse-illusion views in this respect.[24]

There is a core of truth in PIH. My own view upholds its key idea: we represent phenomenal properties as primitive. Still, PIH cannot be the whole story about the intuition of irreducibility. It explains our shared, compelling, and conceptual irreducibility intuition about the phenomenal, but not why this intuition *keeps bothering us*.

[24] For a comparable diagnosis, see Chalmers (2018, 25–29).

Let me explain. Plausibly, we represent external *colors*—properties of surfaces, not of experiences—as primitive. So, we intuit them to be irreducible. Still, the intuition does not bother us. We readily admit that primitive Edenic colors do not really exist and simply seem to exist. Only ordinary, non-primitive colors exist. We have shared, compelling, and conceptual irreducibility intuitions about Edenic primitive colors, but we set them aside as irrelevant—as merely concerning non-existent properties. With some exceptions discussed later, few people think there is a hard problem of colors.

In contrast, our irreducibility intuition about phenomenal properties *does* bother us quite a lot. We are extremely reluctant to set it aside as irrelevant and bearing on mere non-existent properties. We struggle to admit, and even to represent to ourselves, that these primitive phenomenal properties do not really exist and simply seem to exist. This must be explained to fully account for our intuition of irreducibility.

4.4. The illusion meta-problem

A. The illusion meta-problem and the intuition of obviousness

We must explain why it is so hard to admit and even to represent to ourselves that phenomenal properties do not really exist, but simply seem to exist. We must solve what I call the *illusion meta-problem*: explain why the illusory nature of phenomenal consciousness is so hard to grasp, and why, as a result, the illusion of phenomenal consciousness is uniquely compelling. If we solve the illusion meta-problem, we explain why our irreducibility intuition about the phenomenal keeps bothering us, contrary to irreducibility intuitions about external colors. Solving this problem, we understand why illusionism is so hard to believe and even contemplate as such: why many find it not just counterintuitive, but "crazy," "silly," "incoherent," etc.

To solve the illusion meta-problem, it seems natural to first explain our intuition of obviousness about the phenomenal. Remember, this intuition has two components. We intuit phenomenal experiences to be self-revealing: such that, if someone has a phenomenal experience, the whole nature of this experience's phenomenal properties introspectively appears to the subject. We also intuit them to be infallibly appearing in introspection: such that, if it introspectively appears to someone that they have an

experience with a phenomenal property, they really have an experience with this property.

Why is this intuition of obviousness related to the illusion meta-problem? Suppose you encounter illusionism for the first time. You hear that phenomenal consciousness is an illusion and that the phenomenal aspect of the experience you introspect at the moment—say, your visual experience of a white wall—does not exist, but introspectively seems to exist. You try to represent to yourself what this means. Doing so, it occurs to you that, if this phenomenal experience of white introspectively seems to you to exist, it *must* exist. You conclude that what the illusionist says just cannot be right.

Plausibly, this is why illusionism is so hard to represent to ourselves. But a thought of this sort seems natural precisely when we think of phenomenal consciousness as introspectively obvious—as something about which one cannot draw an appearance/reality distinction. When asked to envision that various non-phenomenal entities—a table, our hands, the external world—do not exist, but seem to exist, we find it implausible. However, we still contemplate these possibilities without difficulty. We see nothing incoherent or absurd in the idea of a table, a hand, or a whole external world that does not exist, but seems to exist. It is only for phenomenal experiences that the possibility seems incoherent, because we conceive of them as introspectively obvious.

We should focus on explaining the intuition of obviousness. First, it is, anyway, a part of our explanandum. Second, as I showed, explaining it helps us with the illusion meta-problem. Third, doing so would help us explain why the intuition of irreducibility about the phenomenal keeps bothering us, since the issue is that, contrary to what happens with Edenic colors, we cannot set it aside as irrelevant and concerning merely apparent, nonexistent properties.

The *illusion meta-problem* is not the same as Chalmers's *meta-problem of consciousness*—the problem of explaining mystery intuitions. The illusion meta-problem is a *part* of the meta-problem of consciousness and also a part of the illusion problem. To avoid confusion, Chalmers referred to my illusion meta-problem as the "resistance problem,"[25] but this misleadingly suggests that it is about explaining a doxastic resistance, while I think it is about explaining a representational difficulty, which is upstream of a doxastic resistance. So, I stick to "illusion meta-problem."

[25] Chalmers (2018, 29).

B. AST, QIH, and the illusion meta-problem

AST and QIH might explain our reluctance to believe illusionism by appealing to the robust nature of our representations of phenomenal experiences: whatever we otherwise believe, we keep representing that we have phenomenal experiences. This generates a strong tendency to believe that we have them, and we resist illusionism. However, there are many things we resist believing: e.g., I resist the view that I do not have two hands, because perception strongly and robustly disposes me to believe that I do. What is special about illusionism about phenomenal consciousness is not just that it is hard to believe, but that it is hard to represent to ourselves. AST and QIH do not have much to say about this. At best, they explain a doxastic resistance, not the representational difficulty upstream of the resistance.

Modeling the illusion of phenomenal consciousness on other illusions, we struggle to solve the illusion meta-problem. This is clear if we use the model of cognitively penetrable illusions involving fallacies, like theoretical eliminativists do. However, we also struggle to do so if, like new-wave illusionists, we use the model of robust and cognitively impenetrable illusions—perceptual illusions, such as the Müller-Lyer illusion.[26] If the illusion of phenomenal consciousness were like a perceptual illusion, it would be compelling, and illusionism would be hard to believe, but it is not clear that the illusion would be uniquely compelling and difficult to overcome. It is also not clear that illusionism would be hard to represent to ourselves. To solve the illusion meta-problem, we must go deeper. We must leave behind both the mistakist and the perceptualist stances on the illusion of phenomenal consciousness, and understand what makes this illusion exceptional.

C. The limits of introspection

QIH alone does not help much with the illusion meta-problem, as Pereboom notices. He wonders why "we are resistant to the possibility of [introspective] qualitative inaccuracy" even if we are not as reluctant to the possibility of *sensory* inaccuracy.[27] To explain this, he formulates hypotheses going beyond QIH. He stresses that, with sensory representations, "we have readily

[26] Frankish (2016, 18).
[27] Pereboom (2011, 23–24).

available ways of checking the entity represented that are independent of the representation under scrutiny," while "such ways of checking are at best very limited" with introspection. For Pereboom, this explains our reaction to illusionism: since we have, at most, rare opportunities to notice a discrepancy between introspective representations of phenomenal experiences and the actual experiences themselves, we conclude that these discrepancies do not exist and that introspective appearances of phenomenal experiences are always accurate. This makes us resist illusionism.[28]

In the same spirit, we could also stress that introspection has other features distinguishing it from sensory perception.[29] We have less antecedent knowledge about its functioning, and no strong antecedent belief that it detects standard physical features. This might be what prevents us from admitting the possibility of fallacious introspective appearances.

None of this helps. The idea behind these considerations is roughly that we inductively infer the inexistence (or the impossibility) of introspective inaccuracies from the absence (or rarity) of observed such inaccuracies. This only predicts an inferential and *cognitively penetrable* tendency to deny introspective inaccuracies, but our difficulty in representing to ourselves the illusory nature of the phenomenal is robust and cognitively impenetrable.

A thought experiment makes it clear.[30] Suppose Earth is visited by scientifically advanced aliens from Deneb. Denebians tell us they possess a scientific theory—the U theory—better than our physics, and so different from it that it cannot correctly be called a "physical" theory. Regarding the theory, Denebians only reveal that it posits spatio-temporally localized entities—*delta entities*, about which we know nothing else. Denebians return to their planet and leave behind a delta entity *detector*. We can turn it on and know when it indicates a delta entity, even if we ignore how it works.

Imagine human physicists using the detector to reconstruct the U theory. They have to trust the detector somehow, which is their only means of accessing delta entities. Still, would they be able to *envision* and *represent to themselves* the possibility that the detector is inaccurate? Obviously, yes. They would even often fear this inaccuracy, which would doom their effort. Even if they fully believe that the U theory is correct, and even if they think the detector is accurate, they could easily envision that it dysfunctions.

[28] Others have accepted this explanation of our intuition of introspective infallibility (Shabasson 2022, 432–33). My criticism of Pereboom applies to these views.
[29] Joseph Levine suggested these additions.
[30] Kammerer (2018a).

About this imaginary detector: (1) We have no independent way to check on the accuracy of its measures. (2) We have no antecedent knowledge of its functioning. (3) We have no antecedent belief that it detects standard physical features. Yet, we could easily represent to ourselves the possibility that it is inaccurate. So, the fact that introspection satisfies (1), (2), and (3) cannot explain the fact that we struggle to represent it as possibly inaccurate.

D. Introspection and imaginative conceptions

Pereboom granted that this is a problem.[31] In response, he noted that it would be easy for us to form an imaginative conception of an inaccurate Denebian detector: we could "[extrapolate] from our own experience with detection technology" and "imagine other devices that would serve this checking function and yield different readings." However, we cannot do this with introspection: we "can't easily imagine, for instance, another mind accessing my phenomenal states and representing them differently from how I introspect them." Pereboom's idea is that we cannot represent discrepancies between introspection and phenomenal reality because we lack an imaginative conception of these discrepancies. Why do we lack it? For Pereboom, again, because we have few ways of checking the accuracy of our introspective representations, and thus few opportunities to notice discrepancies, which hinders our ability to imagine them.

To explain our lack of imaginative conception of introspective discrepancies, Pereboom adds a hypothesis. Suppose that, as a matter of fact, introspective representations of phenomenal experiences are themselves introspected as phenomenally conscious. To use the detector analogy, it is as if we had a multitude of delta entity detectors, but that "regardless of whether a detector indicates delta entities, testing its detection display invariably indicates them." If we only represented these detectors through self-directed detection, we would always represent these detectors as themselves containing delta entities. If these representations constrain our imaginative capacities, we always imagine these detectors as *containing delta entities*. Pereboom thinks that this happens with introspection: we always introspectively represent our introspective representations of phenomenal experiences

[31] Pereboom (2019, 191).

as themselves phenomenal. This constrains our imaginative conception, making us always think of them as phenomenal. As a result, we struggle to form an imaginative conception of what illusionism depicts: representations of phenomenal experiences without phenomenal experiences.[32]

Pereboom is on the right track when he insists on our difficulty in forming an imaginative conception of illusionism. This brings us closer to the truth: the issue is not primarily a matter of doxastic resistance, but of representational difficulty. Still, this explanation also fails. First, it fails to distinguish between de facto and de jure unimaginability. Second, what needs explaining here is not just a lack of imaginative conception, but more generally a lack of positive conceptions.

Pereboom's starting point is that we never, or rarely, represent certain things. We rarely represent discrepancies between introspection and phenomenal reality. We never introspectively represent a non-phenomenal introspective representation of phenomenal states. As we never (or rarely) represent these things, we cannot imagine them. However, this explains a de facto unimaginability, not a de jure one. Here is a way to approach the distinction. Suppose I have never tasted durian. I cannot imagine its taste, even if it were described to me. However, this inability is de facto. I know that the taste of a durian is something imaginable, even if I cannot make the right image available to me. Now, suppose someone asks me to imagine a round square, or a face looking exactly like my mother's *and* Donald Trump's, or a surface entirely red and green. I cannot imagine those things. Contrary to the taste of the durian, they are unimaginable in principle, de jure.

When an imaginative inability stems only from a lack of exposure to relevant representations, plausibly, the unimaginability is only de facto. Pereboom's explanation of our inability to imagine certain introspective discrepancies appeals to such a lack of exposure to relevant representations. Therefore, the unimaginability of illusionism should be de facto. However, trying to represent what illusionism describes, many feel they are grappling with something "crazy," "silly," "absurd," "incoherent," etc. It is not just that they lack the right mental image. Illusionism seems to be something

[32] This second hypothesis explains (a) an inability to imagine introspective representations of phenomenal states *without any phenomenal state being present*, not (b) an inability to imagine *inaccurate* introspective representations of phenomenal states. Explaining (a) without (b) helps solve the illusion meta-problem, but does not explain the intuition of obviousness in full.

unimaginable in principle, somewhat like a surface entirely red and green, or a face that looks exactly like Trump's and my mother's at the same time. Illusionism seems hard to represent de jure.

Pereboom could retort that, about consciousness, we mistake a de facto unimaginability for a de jure one. I doubt this is the case, but for the sake of the argument, suppose this is correct. Still, a second issue arises. Pereboom's explanation focuses on failures to form imaginative conceptions, but what is at stake here is a more general difficulty in forming positive conceptions.

Consider a variation on the Denebian detector thought experiment, involving an *Unimaginable Detector*. In this scenario, Denebians never show us detectors of delta entities. They tell us these detectors exist, but cannot be perceived or imagined. Now, suppose they inform us that, at time t, detector D indicates the presence of a delta entity. We know we cannot imagine D. Still, can we coherently represent to ourselves the possibility that D is inaccurate and that it misrepresents the presence of a delta entity? Obviously, yes. This is not difficult to represent, although not through specific imagistic representations. We can positively conceive of many things we cannot imagistically represent—we know this at least since Descartes's chiliagon. So, an inability to form an imaginative conception of some representational discrepancy is not sufficient to explain an inability to positively conceive of this discrepancy. There are many things that we cannot imagine in the strict sense, that we nevertheless conceive of, without finding them absurd or incoherent in the slightest. However, when people struggle to represent to themselves the truth of illusionism, the view strikes them as crazy, absurd, and incoherent. It is hard for them to grasp, not just to represent imagistically. Pereboom's hypothesis, with its focus on imaginative conception, remains unable to explain this difficulty. It does not solve the illusion meta-problem.

E. Conclusion on the illusion meta-problem

All these views fail to solve the illusion meta-problem. They do not explain why we find illusionism so hard to represent to ourselves. They do not explain our intuition of obviousness. To move forward, we need a view that does not model the illusion of phenomenal consciousness on reasoning mistakes (mistakism) or perceptual illusions (perceptualism), but acknowledges its unique character.

4.5. The evidential approach

I propose a strategy I call the *evidential approach*.[33] Its key idea is to understand how our (illusory) representations of phenomenal experiences are intertwined with the cognitive processes we use to represent *evidential* situations—what appears to us, what is epistemically given to us, and what counts as the evidence at our disposal. *Phenomenal* and *evidential* cognition are deeply entangled.

This strategy recently attracted attention. David Chalmers, associating it with the work of Wolfgang Schwarz and with my own work, suggested that it is "the sort of idea we should be exploring."[34]

On a first approach, why follow this strategy? The intuition of obviousness is the intuition that phenomenal experiences are *presented* to the conscious subject—that they *appear* to the subject—in a peculiar way: that they are *given* to the subject in a specifically intimate manner. Having said so, we notice that the concepts used to articulate the content of this intuition are epistemological and evidential concepts: concepts of presentation, appearance, given-ness, obviousness, etc. It thus seems promising to posit a relation between our representations of phenomenal properties (phenomenal cognition) and our representations of evidence, appearances, presentations, given-ness, or obviousness (evidential cognition).

A. Introspective opacity

The idea that mystery intuitions are linked to epistemological specificities of our grasp of consciousness is not new. Some canonical views take this path. They share some of the spirit of the evidential approach, broadly conceived.

For instance, consider views with similarities to Graziano's AST, which focus on introspective opacity.[35] For these views, introspection is opaque: it tells us *which* states we are in, but does not reveal their details, nor the details of the introspective process—like Graziano's attention schema, which does not reveal the details of the attentional processes it represents. This supposedly generates a sense that introspection gives us *immediate*

[33] For the expression, see Kammerer (2019c).
[34] Chalmers (2020b, 224).
[35] There are many such views (Armstrong 1968; Chalmers 2018, 22–24; A. Clark 2000; Dennett 1991).

knowledge: introspecting my experience of red with no access to the causal or inferential mediation leading to my introspection, I falsely conclude that I know about my experience without mediation.

Even if an intuition of immediate knowledge were sufficient for an intuition of obviousness (it is not: something could be the object of immediate but fallible and non-revelatory knowledge), this explanation would be lacking. These appeals to introspective opacity predict that the resulting intuition should be *cognitively penetrable*, since it is generated by an *inference* from the opaque deliveries of introspection. However, our intuition of obviousness is not cognitively penetrable.

Counterexamples also show the limits of these views. We have many beliefs whose causal or inferential history we ignore. I believe that my name is François, but I am unaware of causal or inferential mediations leading me to this belief. I *suppose* there is a causal chain that led to this belief, starting with my parents' decisions, but I am not aware of its components. Still, my name does not seem at all obvious to me as phenomenal experiences do. Intuitively, I see how I could have a name without being aware of it, or how I could be wrong about the name I think is mine.

B. Perceptual Reality Monitoring

Other views are close to the evidential approach in different respects. Take the "Perceptual Reality Monitoring" (PRM) theory,[36] a theory of consciousness that belongs to the family of higher-order theories (more on those in Chapter 8). For PRM, a representation becomes phenomenally conscious when appropriately meta-represented by a monitoring mechanism ("discriminator") which determines whether representations faithfully depict reality, and therefore, whether beliefs based on these representations are "justified."[37]

PRM has an affinity with the evidential approach: it gives a central explanatory role to evidential representations. A key difference is that PRM and the evidential approach have different explananda. The evidential approach aims to explain mystery intuitions, and notably the intuition of obviousness. PRM aims at explaining *conscious experience* itself, not our mystery

[36] Lau (2022); Michel (forthcoming).
[37] Lau (2022, 157).

intuitions. In fact, PRM's main defenders think our mystery intuitions are *not* compelling and are primarily caused by sociological factors.[38] It is thus unsurprising that they do not try to explain them by appealing primarily to the widely shared cognitive mechanisms they posit to explain consciousness itself.

C. Designed certainty views

Some recent views come even closer to the evidential approach as I see it. I call them "designed certainty" views.[39] Their key idea is that, as our cognitive systems (as Bayesian engines) must update their beliefs upon incoming sensory input, it can be advantageous to entertain a sort of certainty toward some propositions, in order to help with belief updating.[40] These propositions we hold with certainty are roughly those we could believe based on phenomenal introspection. Call corresponding beliefs "phenomenal beliefs."

A nice feature of designed certainty views is that they view the epistemological peculiarity of phenomenal introspection as a robust feature, not a bug. However, they fail to explain our intuition of obviousness.

Holding phenomenal beliefs with certainty and having an intuition of obviousness about the phenomenal are two distinct things. The first is a putative feature of our phenomenal beliefs, corresponding to the fact that we hold them with *complete* (or near-complete) confidence. The second is a shared, compelling, and conceptual disposition to believe that phenomenal experiences are introspectively obvious: it is a disposition to form certain *meta-introspective* beliefs. These meta-introspective beliefs are not phenomenal beliefs. They are not about having phenomenal experiences, but about the relation between introspection and phenomenal experiences.

We should not overlook this difference. The polysemy of "certainty" is misleading. It has a psychological sense: a belief is psychologically certain when held with full confidence. It also has an epistemological sense: a belief is epistemologically certain when it is maximally justified and can be justifiably held with full confidence. In the epistemological sense, but not the

[38] Lau and Michel (2019).
[39] For the expression and a first version of the following criticism, see Kammerer (2023).
[40] Wolfgang Schwarz first developed this sort of view (Schwarz 2018; 2019), later taken up and modified by others (A. Clark et al. 2019).

psychological sense, certainty and the intuition of obviousness are tightly linked: our tendency to believe that phenomenal states are obvious, plausibly, is close to a disposition to believe that introspective phenomenal beliefs are epistemologically certain. However, designed certainty views merely hypothesize that phenomenal beliefs are *psychologically* certain.

Here is another way to make the difference striking. If what happens is just that we are 100% confident of our phenomenal experiences, we should reject illusionism with 100% confidence, but illusionism should not be hard to represent to ourselves. We could still grasp what it depicts clearly and coherently, while thinking with full confidence that it is false.

So, the certainty of phenomenal beliefs and our intuition of obviousness are distinct features. Could there be a *causal* relation between the two? If the first causes the second, it explains it: designed certainty views explain the intuition of obviousness.

However, it is unclear why the certainty of phenomenal beliefs should cause an intuition of obviousness. Arguably, we hold some non-phenomenal beliefs with certainty: e.g., logical beliefs ("If P and, if P, then Q, then Q"), definitional beliefs ("All bachelors are unmarried"), or mathematical beliefs ("2 + 2 = 4"). Still, we do not intuit the related entities as obvious. They do not seem self-revealing: e.g., there is an infinity of numbers never presented to us. They do not seem infallibly appearing in the same way as phenomenal experiences: Descartes's cogito famously stressed that an evil demon could provide us with a fallacious appearance of mathematical truths, but not with a fallacious appearance of conscious thoughts. This suggests that we do not have an intuition of obviousness about mathematical objects, even if mathematical beliefs are psychologically certain. However, if the certainty of beliefs fails to generate an intuition of obviousness in some cases, why would it do it in the phenomenal case?

These counterexamples concern beliefs established a priori. Some other, more debatable counterexamples concern a posteriori beliefs, which seem held with certainty without any intuition of obviousness. Consider subjects victims of monothematic delusions. On some plausible views, they hold "strange" beliefs with certainty: nothing can make them change their minds. Yet, they do not develop an intuition of obviousness about the corresponding entities. Capgras subjects think a nefarious lookalike has replaced a loved one. Cotard subjects think they are dead or non-existent. All apparently hold these beliefs with full confidence. Still, when asked *why* they believe these things, they do not mention a unique access to obvious facts, but only

mundane and insufficient evidence. In some cases, they concede that they do not have good reasons to have the beliefs they have. So, even a posteriori certainty seems insufficient to generate an intuition of obviousness.[41]

The evidential approach is a promising way to explain the intuition of obviousness and solve the illusion meta-problem. However, existing views falling close to this approach fail.[42]

* * *

I presented the illusion problem, as well as promising candidate solutions sharing a *Common Core* I intend to uphold. However, these promising views fail to solve the problem, notably because they fail to explain our intuition of obviousness and do not solve the illusion *meta*-problem. To move forward, I introduced the evidential approach. I showed that existing views close to this approach also fail. I turn to my own view, which is a version of the evidential approach: Meta-Cartesianism.

[41] For more details, see Kammerer (2023, 14–19). Of course, some theorists deny that delusions count as beliefs.

[42] Other views, arguably also close to this approach, cannot be discussed in detail here, but I think they also fail for various reasons. I mentioned earlier the view of Shabasson (2022), which I think fails for the same reason as Pereboom's. There is also the view of Murez and Thouzeau-Corpel (2024), on which phenomenal concepts satisfy a "super-transparency" constraint, making us reject identifications of their referents with those of any other concepts—a view I fear makes wrong predictions about irreducibility intuitions.

5

Mapping the House of Mirrors

Introducing Meta-Cartesianism

Consciousness: late, poorly developed, for external purposes, exposed to the grossest errors, and even essentially something falsifying, coarsening, simplifying [. . .]. The external phenomenology gives us, by far, the richest material, and allows the most rigorous observation; while the internal phenomena are difficult to grasp and more prone to errors.

—Friedrich Nietzsche

5.1. Meta-Cartesianism

A. Meta-Cartesianism in a nutshell

I will now present my candidate solution to the illusion problem: Meta-Cartesianism. For Meta-Cartesianism, we represent phenomenal experiences as entities playing a peculiar evidential role and standing in a uniquely intimate epistemic relation to us. Yet, there are no such entities, no phenomenal experiences. When we represent our phenomenal experiences, we misrepresent what happens in our minds. Why does this misrepresentation occur? Because it is a convenient way to make sense of the evidential contribution of, notably, sensory processes, which we acquired through evolution.

We represent phenomenal experiences as entities with which we have a uniquely intimate epistemic relation. As a result, we cannot intuitively grasp their illusory nature: about them, we cannot draw a certain sort of appearance/reality distinction. We represent the phenomenal as *obvious*. This is why we struggle to represent illusionism about the phenomenal, even if, correctly understood, it is coherent and true. This makes the illusion of phenomenal consciousness exceptional.

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0006

Meta-Cartesianism solves the illusion problem. It succeeds where other views fail. It follows the evidential approach, positing the entanglement of phenomenal and evidential cognition. It is both *meta*-Cartesian and *anti*-Cartesian: for Meta-Cartesianism, Descartes was wrong about what our mind really is, but his view was a roughly accurate transcript of how we introspectively represent it.

Meta-Cartesianism is the last iteration of a set of ideas I developed in the last decade under various names. Its ancestors were called the "Theoretical Introspection Hypothesis" in 2016,[1] and the "Theoretically determined Concepts of Epistemologically special states" theory in 2021.[2] There are some differences between these different versions of the same view, and I briefly mention some of them at the end of the chapter. The name changes show my own indecisiveness in the matter; hopefully, "Meta-Cartesianism" will stick.[3]

B. Meta-Cartesianism's three tenets

The Meta-Cartesian view has two main posits: the *Cartesian Device* and the *Evidence-by-Resemblance Mechanism* (ERM). It has three related central tenets: *Architecture, Implementation*, and *Illusionism*. I lay them down here and examine them further below.

Architecture: We use a Cartesian Device for evidential and phenomenal cognition.

Implementation: We implement a Cartesian Device with an Evidence-by-Resemblance Mechanism.

Illusionism: There are no phenomenal experiences: our Cartesian Device's representations of phenomenal experiences are misrepresentations.

We represent phenomenal experiences with a Cartesian Device. A Cartesian Device is an innate module used in evidential cognition and phenomenal cognition. It deploys phenomenal concepts, which track certain mental processes, like sensory processes, and represent them as *phenomenal experiences*.

[1] Kammerer (2016c).
[2] Kammerer (2021). For other iterations, see Kammerer (2016a; 2016b; 2019a).
[3] Many thanks to David Chalmers for suggesting the name.

Phenomenal experiences are represented as primitive states playing a peculiar evidential role. This role is as follows. First, phenomenal experiences provide us with *direct evidence* for certain things: they provide us with non-inferential justification, which is independent of what we otherwise believe. Second, and crucially, they also stand themselves in a uniquely intimate evidential relation to us: they are *self-revealing* and *infallibly directly presented*.

The idea that we use a Cartesian Device to represent phenomenal experiences, grasped as specific evidential states, is summarized in *Architecture*, the first tenet of the Meta-Cartesian view:

> *Architecture*: We use a Cartesian Device for evidential and phenomenal cognition.

How does our Cartesian Device function? How do its representations gain their content? For Meta-Cartesianism, it relies on an ERM. That is: our Cartesian Device recruits a rough measure of resemblance between items to assess the direct evidence our experiences provide. It reckons that phenomenal experiences present to us what they distinctively resemble. Given plausible assumptions that will be presented later, this leads to representing phenomenal experiences as providing direct evidence *and* as being epistemically given to us in a uniquely intimate way.

This idea is captured in Meta-Cartesianism's second tenet:

> *Implementation*: We implement a Cartesian Device with an Evidence-by-Resemblance Mechanism.

Finally, Meta-Cartesianism is *meta-Cartesian*, but also *anti-Cartesian*: our Cartesian Device's representations are *misrepresentations* of the real phenomena they track. They are not satisfied by anything real: there are no phenomenal experiences. This is Meta-Cartesianism's third tenet:

> *Illusionism*: There are no phenomenal experiences: our Cartesian Device's representations of phenomenal experiences are misrepresentations.

I will now detail the view and make a preliminary case for it. First, adopting the design stance, I give a partial argument for *Architecture* and show that a Cartesian Device would be convenient for evidential cognition. Then, I give a preliminary argument for *Implementation*: I show that an

Evidence-by-Resemblance Mechanism is a plausible way to implement a Cartesian Device. I then combine these two theses with illusionism. Finally, I clarify Meta-Cartesianism by comparing it to other views. In the next chapter, I show that it solves the illusion problem.

5.2. Evidential cognition from the design stance

In this section and the next, I argue for the following. If we have to design a cognitive agent,[4] it makes sense to endow it with evidential cognition. It also makes sense for this evidential cognition to rely on a Cartesian Device: a modular device representing certain real mental processes, like sensory processes, as primitive states of direct evidence, themselves epistemically given to the agent in a uniquely intimate way—evidentially self-revealing and infallibly presented. Therefore, there is some plausibility to the view that our evidential cognition relies on a Cartesian Device. This makes a partial case for *Architecture*.

A. René and his limits

Consider René, an imaginary humanoid robot. He has sensory access to the world through sensory captors and sensory modules, which have an appreciable degree of informational encapsulation, automaticity, and central inaccessibility. He has metacognitive representations of cognitive agents, beliefs, desires, and so on, which he applies to others in mindreading and to himself in introspection, thanks to a metacognitive module: René is self-conscious.

The output of René's sensory modules is formulated with *sensory* concepts, tracking real complex phenomena (e.g., reflectance properties of surfaces), conveniently mischaracterizing them as primitive sensory qualities (e.g., Edenic colors), while implicitly placing them in quality spaces corresponding to the set of complex discriminations and classifications enabled by corresponding sensory capacities.

The output of René's metacognitive module is formulated using metacognitive concepts, tracking complex functional internal states, conveniently

[4] For other approaches taking the design stance, see Chalmers (1996, 184–86; 2018, 20); Dennett (2019, 56–57); Graziano (2019b). On reverse engineering, see Dennett (1995b); Miłkowski (2013).

mischaracterizing them as primitive relations (e.g., a belief being a primitive relation between a subject and a proposition), which are also implicitly apprehended as playing certain functional roles.

René has imagination, memory, and a basic innate repertoire of abstract concepts. He has a global workspace that gathers the modules' outputs and makes them accessible for the formation of central beliefs, planning, decision-making, reasoning, and so on: René is access-conscious. He is embedded in a physical and social environment, and has peers (Elisabeth, Blaise, Antoine, etc.), with whom he communicates and cooperates.

René lacks two related things. First, he is not (yet) under the illusion of phenomenal consciousness. He has no phenomenal cognition and no phenomenal concepts. He introspects his mental states as primitive states, but, as I argued in the previous chapter when criticizing Pereboom's PIH, this is not enough to explain the appearance of phenomenal consciousness, and notably the appearance of *obvious* states.

Second, he lacks evidential cognition. He has no concept of evidence. The concept of evidence I have in mind is a mentalistic personal-level concept representing a pro tanto *reason to believe a proposition*, provided by a mental state; a reason which weighs in favor of believing this proposition.[5] This concept is normative. It concerns what one should believe. It allows evidence to be either deceptive or trustworthy. This concept is roughly equivalent to a concept of propositional justification or epistemic appearance. Lacking evidential cognition, René can think of what is but not of what epistemically appears to be. He can think of what he *believes*, not of what he should believe.

B. The advantages of evidential cognition

Pre-enhancement René lacks evidential cognition. From a design stance, it seems a good idea to give it to him. Why?

First, evidential cognition improves René's belief-formation: he can *scrutinize* his beliefs and represent his evidence for them, potentially realizing his beliefs do not match his evidence. Sophisticated belief-formation can occur without personal-level explicit evidential cognition, but evidential cognition helps: e.g., explicitly reasoning about evidence allows René to understand a

[5] In what follows, evidence is always evidence *for a proposition*, but for reasons of convenience, I will talk of evidence for a *belief*—i.e., evidence for a proposition complementing a given possible belief—or a *state of affairs*—i.e., evidence for a proposition made true by a given state of affairs.

class of states as deceptive (say, visual experiences had in abnormal light) and update his beliefs accordingly.

Second, representing the evidence of others, René better understands and predicts their behavior. Suppose he knows a venomous snake is in front of him, but understands that Blaise has visual evidence the snake is a rock—it looks like a rock to him. René can predict that Blaise will believe he faces a rock and kick the snake instead of avoiding it.[6]

Third, once represented, evidence is communicable. This fosters epistemic cooperation.[7] If René tells Elisabeth he has belief B, she might form B too. If he tells her he has B based on evidence E, which Elisabeth knows to be deceptive, she avoids forming B and helps René back by sharing her assessment of E.

Fourth, with evidential cognition, René better assesses epistemic agents. He might see that Elisabeth is better than Blaise at *following* evidence, while Blaise is better at *gathering* evidence (Elisabeth may be sharp-minded but short-sighted). He can then trust them for different things. René can also better self-assess and self-improve.

5.3. Designing evidential cognition

How should we design René's evidential cognition? I will argue that it would make sense to give him a Cartesian Device.

A. Representing evidence

To acquire evidential cognition, René first needs a representation of evidence. Pre-enhancement René represents agents, beliefs, and mental states. Plausibly, in his innate conceptual repertoire of abstract concepts, there is also a normative representation of *obligation* or *to-be-done-ness*—the concept SHOULD.

If so, René can combine his existing representations into a representation of evidence: a representation of a mental state such that, in virtue of

[6] For the idea that representing perceptual evidence helps predicting others' behavior in illusory settings, focusing on animals, see Lurz (2011, 83–84).

[7] For the benefits of epistemic cooperation in perceptual tasks, see Bahrami et al. (2010); Bang et al. (2014).

this state, an agent should believe something to a certain extent. The combination itself can remain inaccessible to René and be only manifested in the representation's functional role.

B. Direct and indirect evidence

René then deploys this concept to represent evidence—his and others'. I focus on his own evidence here.

Three things seem plausible. First, it makes sense to have representations of evidence track *reliable* belief-formation processes. A *belief-formation process* is a process upstream of beliefs, which *tends* to generate or change beliefs, even if it does not always do so. A belief-formation process is reliable if, when it leads to a belief, it is usually a true belief. Saying that a belief-formation process is reliable, in my mind, is conceptually different from saying that it provides evidence. Still, *tracking* reliable belief-formation processes and *representing* them as evidence makes sense, from the design stance, to promote epistemic success.

Second, René, with its modular architecture, has two broad sorts of belief-formation processes. On the one hand, *inferential* processes. For instance, if René believes that P and that, if P, then Q, he can infer that Q. On the other hand, *non-inferential processes*: typically, endorsements of modular outputs that have been globally broadcast. For instance, when his visual module's output that there is something red is broadcast in his global workspace, René often concludes that there is something red, without inferring it from his previous beliefs. Of course, modular processes involve internal computations that count as inferential in a sense, but here I reserve "inferential" for personal-level inferences based on central beliefs: endorsing the output of an encapsulated module is non-inferential.

Third, it makes sense to design representations of evidence so that they track and characterize differently distinct ways of forming beliefs.

On these three assumptions, it is natural to design René to distinguish inferential and non-inferential belief-formation processes and represent them as providing different sorts of evidence.

Good inferential processes are conditionally reliable: from true beliefs, the proper inferences tend to yield other true beliefs. It makes sense for René to represent them as *indirect evidence*: evidence essentially dependent on other beliefs of the agent. Good non-inferential processes are reliable

independently of one's beliefs. It makes sense for René to represent them as *direct evidence*: evidence essentially independent from the beliefs of the agent. Intuitively, representing one's indirect evidence answers: "Given what I already believe, what else do I have reasons to believe?". Representing one's direct evidence answers: "Independently of what I already believe, what do I have reasons to believe?"

C. Direct evidence is modularly represented as primitive

As indirect evidence depends on one's beliefs, René's representations of his indirect evidence should be sensitive to his beliefs. He should represent indirect evidence through central and informationally promiscuous processes. By contrast, direct evidence is independent from one's beliefs: it makes sense for René to represent it in a manner insensitive to his beliefs, via an encapsulated module.

It would be natural for this module to be an extension of René's metacognitive module. Representations of direct evidence would be similar to other metacognitive representations and characterize their referents as *primitive*. Direct evidence could be represented as a primitive relation between a subject and the supported proposition, in virtue of which the subject has a pro tanto reason to believe the proposition, independently of what the subject believes.

To summarize, René would track differently inferential and non-inferential belief-formation processes. He would represent these, respectively, as indirect and direct evidence. Direct evidence would be represented via an encapsulated module and characterized as primitive. Finally, while the focus was on René's representations of his own evidence, it makes sense for him to deploy similar representations about others' evidence.

D. Direct evidence as self-presenting

René has beliefs about *his own direct evidence*. Call these beliefs "evidential self-beliefs." How does he represent evidence supporting them?

First, consider indirect evidence. René believes he has *indirect evidence* for the belief that Q, and that this indirect evidence depends on his belief that P and his belief that, if P, then Q. It seems convenient for him to represent

that the evidence for his evidential self-belief is no other than his evidence for P and his evidence for the proposition that if P, then Q—as well as, say, his evidence for the inference's legitimacy. If so, René can represent chains of indirect evidence ("Why believe that P? Because I believe that Q. Why believe that Q? Because I believe that R, etc."), tracking chains of potential inferences.

Second, consider direct evidence. On a convenient design, René represents direct evidence as providing direct evidence for itself: he represents direct evidence as *self-presenting*.

Why? First, he represents direct evidence modularly. His beliefs about his direct evidence are typically caused non-inferentially, by endorsing the output of his evidential module. So, the same evidential module should track its own non-inferential belief-formation process and represent corresponding direct evidence. So, in normal cases, René's evidential module represents that his beliefs about his own direct evidence are supported by direct evidence. (Of course, he sometimes represents that these beliefs are *also* supported by indirect evidence.)

So, the same module represents René's first-order direct evidence (say, evidence that there is something red) and his related second-order direct evidence (say, evidence that he has direct evidence that there is something red). It seems cumbersome to represent these two pieces of evidence, which typically come together, as distinct. On a simpler design, René represents direct evidence as providing direct evidence for its own presence: as *self-presenting*.

Suppose direct evidence is not represented as self-presenting: René has to represent a distinct state of second-order direct evidence as supporting that he has first-order direct evidence, and another distinct state of third-order direct evidence as supporting that he has second-order direct evidence, etc. Grasping direct evidence as self-presenting is simpler. Since René already represents chains of *indirect* evidence, one advantage of grasping direct evidence as self-presenting is that self-presenting direct evidence becomes a natural and convenient stopping point in chains of indirect evidence ("Why believe that Q? Because I believe that R. But why believe that R? Well, R just seems true, independently of any beliefs I have.")

Let us summarize. On a convenient design, René modularly tracks his non-inferential belief-formation processes, like globally broadcast sensory outputs. He represents them as primitive direct evidence, characterized as self-presenting—as giving evidence for its own presence to the subject.

This design is not the only one possible for evidential cognition. On other possible designs, René might take unsupported beliefs as stopping points. Still, this design seems simple and convenient.

E. Direct evidence as self-revealing

René represents direct evidence as primitive: as having a simple and uncomposed nature, captured by his representation of it as direct evidence. He also represents direct evidence as self-presenting: as such that a subject who has some direct evidence also has higher-order direct evidence for the having of this direct evidence. Moreover, he represents that the relevant direct evidence provides the higher-order direct evidence itself: these are not two distinct states.

Together, these features imply that René represents direct evidence as such that, when someone has some direct evidence, they also have ipso facto higher-order direct evidence for this first direct evidence, and this higher-order direct evidence is such that the proposition it supports *captures* the nature of the first direct evidence: indeed, direct evidence being primitive, its nature is captured by a representation of it qua direct evidence.

In other words, René represents direct evidence as *self-revealing*: such that, when someone has some direct evidence, they also have *revealing* higher-order direct evidence that they have this direct evidence: they have higher-order direct evidence for a proposition capturing the nature (or essence) of their first direct evidence. The first direct evidence and the revealing higher-order direct evidence are not two distinct states, but the same state of direct evidence.

F. Direct evidence as intrinsically neutral

René represents direct evidence as primitive. Direct evidence—at least direct evidence for propositions which are not always true or false—can *in principle* be either deceptive or trustworthy. It would be difficult to admit that direct evidence for such propositions has different intrinsic features when it is trustworthy vs. deceptive. Trustworthy and deceptive direct evidence for the same thing intrinsically are the same primitive relation to the same proposition.

Therefore, René represents direct evidence as intrinsically neutral. He implicitly accepts a principle of evidential neutrality: If a subject S has deceptive direct evidence for P in situation C, then, if P is possible, there is a possible 'good' situation C′ where S's direct evidence for P is trustworthy, and the evidence S has for P in C and C′ is intrinsically the same.

G. Direct evidence as infallibly directly presented

As a result, René represents direct evidence as infallibly directly presented: for René, a subject cannot have *deceptive* direct evidence for their own direct evidence.

Usually, René represents evidence as potentially deceptive. When his visual module outputs that there is a red tomato, and the output is broadcast globally, his evidential module detects this non-inferential belief-formation process, and represents it as primitive direct evidence that there is a red tomato. For René, this evidence could be trustworthy (a correct perception of a red tomato) or deceptive (an illusion thereof). As René accepts the principle of evidential neutrality, he represents that, supposing the evidence is deceptive, there could be a case where he has intrinsically identical direct evidence for the presence of a red tomato, but in which the evidence is trustworthy. In other words, René thinks he might be having an illusion of a red tomato, but that this illusion would not wear its illusory nature on its sleeves: an intrinsically identical bit of evidence could be, in other circumstances, a correct perception.

I claimed that René represents his direct evidence as self-revealing: representing that he has direct evidence for the presence of a red tomato, he represents that this piece of evidence counts as revealing higher-order direct evidence for its own presence. The same primitive state by which he has a reason to believe that there is a red tomato also gives him a reason to believe that he has direct evidence that there is a red tomato, in a way that captures the nature of his first-order direct evidence. This sort of higher-order direct evidence is always trustworthy: it *makes itself correct.*

However, for René, could *some* higher-order direct evidence be deceptive? Could he admit, say, of *deceptive* direct evidence that he has direct evidence that there is a red tomato? Direct evidence being self-revealing does not prevent this. It just implies that, if someone really has a given bit of direct evidence, they ipso facto have revealing higher-order direct evidence about it.

Still, it apparently leaves open the possibility that one could have misleading higher-order direct evidence, without the first-order bit.

However, René cannot represent deceptive higher-order direct evidence. To see why, suppose first, as I will show is impossible, that René represents that he has deceptive higher-order direct evidence that he has some first-order direct evidence (e.g., direct evidence that there is a red tomato). For René, direct evidence is intrinsically neutral: there is a possible 'good' situation where his higher-order evidence is intrinsically the same, but trustworthy. In this good case, René actually has the first-order direct evidence.

However, René thinks direct evidence is self-revealing. His first-order direct evidence in a good case supposedly reveals itself: it is itself part of whatever trustworthy higher-order direct evidence he has in a good case. But the deceptive higher-order direct evidence he has in the bad case must be intrinsically identical to the trustworthy second-order direct evidence he has in a good case, which implies that it must also include first-order direct evidence for the presence of a red tomato, making the deceptive higher-order direct evidence correct after all—and trustworthy. The 'bad' case is not in fact bad. René cannot represent higher-order direct evidence as deceptive.

So, as René represents direct evidence as self-revealing and intrinsically neutral, he also represents it as infallibly directly presented: higher-order direct evidence for one's own direct evidence is always trustworthy. This only applies to the direct evidence one has for one's own present direct evidence. René easily represents deceptive direct evidence for *someone else's* direct evidence, or deceptive direct evidence for one's *past* or *future* direct evidence, or deceptive *indirect* evidence for one's present direct evidence.

H. The Cartesian Device

I argued that it would be advantageous for René to have evidential cognition, functioning in a certain way. A mechanism functioning in this way is a Cartesian Device. It is a modular device, possibly part of a metacognitive module, tracking non-inferential belief-formation processes, like broadcast sensory outputs, and representing them as *direct evidence*—evidence one has, irrespective of beliefs. It represents direct evidence as *primitive*. It also represents direct evidence as *self-presenting*, and even *self-revealing*: such that direct evidence always counts as revealing direct evidence for itself. A Cartesian Device also represents direct evidence as *intrinsically neutral*.

As a result, it represents direct evidence as *infallibly directly presented*: such that, when someone has direct evidence that they have some direct evidence, the higher-order direct evidence is always trustworthy.

René and his Cartesian Device are plausible fictions. However, if a Cartesian Device corresponds to a convenient design of evidential cognition, the idea that humans acquired a Cartesian Device through evolution becomes somewhat plausible.

A Cartesian Device tracks globally broadcast sensory outputs and represents them as primitive, self-revealing, and infallibly directly presented direct evidence. The reader sees that this mirrors the intuition of obviousness we have about the phenomenal. Moreover, the fact that a Cartesian Device characterizes direct evidence as primitive suggests that it will generate irreducibility intuitions about it, for reasons detailed in the previous chapter. I develop these ideas at length in the next chapter, but all this already lends some plausibility to the idea that we also use a Cartesian Device for *phenomenal cognition*: our Cartesian Device's representations of direct evidence are our representations of phenomenal experiences, our phenomenal concepts.

So, we now have preliminary reasons to consider *Architecture*, the first tenet of Meta-Cartesianism:

> *Architecture*: We use a Cartesian Device for evidential and phenomenal cognition.

Architecture, at this stage, is just a hypothesis. It will gain support when I show, in the next chapter, that Meta-Cartesianism, of which it is a part, solves the illusion problem.

5.4. Implementing a Cartesian Device

Suppose that proto-humans were analogous to pre-enhancement René. How could they have evolved a Cartesian Device?

A. First steps

First, they needed a concept of evidence. Its components are concepts of beliefs, cognitive agents, and mental states, which proto-humans had, if

they were like René. One last required component is the normative concept SHOULD. Its possession by proto-humans is plausible, given how early in development normative cognition emerges.[8]

Supposing a basic endowment of logical concepts, including a concept of *dependence*, proto-humans could easily acquire the concept DIRECT EVIDENCE. If they were like René, they already represented mental properties and relations as primitive; it seems natural for them to also represent direct evidence as primitive.

Proto-humans then needed to introspectively track non-inferential belief-formation processes, for instance, via an extension of their metacognitive module. Source monitoring theories of metacognition,[9] as well as the Perceptual Reality Monitoring theory of consciousness discussed previously,[10] already hypothesize a "discriminator" mobilizing various heuristics to determine whether representational activities are due to noise, are internally generated, or externally generated. A version of such a discriminator, using slightly different heuristics, could track and distinguish inferential and non-inferential belief-formation processes.

So, suppose proto-humans indeed acquired a module tracking non-inferential belief-formation processes (like broadcast sensory outputs) and representing them as primitive direct evidence. This is not a Cartesian Device yet. What is missing? The characterization of direct evidence as self-revealing and infallibly directly presented. How could it be implemented?

B. The Evidence-by-Resemblance Mechanism

A representation of direct evidence as self-revealing and infallibly directly presented might seem hard to evolve, given its abstract content. Could evolution, tinkering with preexisting cognitive capacities of proto-humans, have landed it?[11] Yes, by building an Evidence-by-Resemblance Mechanism (ERM).

Here is how it works. Suppose a sensory module outputs THERE IS RED. A discriminator detects that this output is a non-inferential belief-formation

[8] Three-month old showing preferences for pro-social behavior might already grasp actions as actions that should be done (Hamlin et al. 2010). On basic normative cognition in children and animals, see Schmidt and Rakoczy (2023).

[9] Lindsay (2008).

[10] Lau (2022).

[11] For evolution as a tinkerer, see Jacob (1977).

process. It feeds it to the ERM. The ERM redeploys and reformulates the output THERE IS RED, using a conceptual operator (*...*), to produce a metacognitive output: THERE IS *RED* IN ME. While RED is a sensory concept, *RED* is a metacognitive concept. Both characterize their referent as primitive. They almost share their representational vehicle, save for the operator *...*.

This operator *...* has tight functional links to the concept DIRECT EVIDENCE. Concepts formed with the operator conform to the following rule:

- *Evidence-by-Resemblance Rule*: *A* provides direct evidence, to the subject in *A*, for what *A* distinctively resembles.

"Distinctively resembles" means the following: *A* distinctively resembles B when *A* resembles B and *A* is, among all *x* states, the one resembling B the most, or equally the most.[12]

The *Evidence-by-Resemblance Rule* invokes resemblance, but the ERM does not explicitly represent resemblance. It just manipulates concepts formed with the operator *...*, like *RED*, in a way conforming to this rule. It mobilizes an *implicit* measure of resemblance between items to determine the direct evidence provided by *x* states.

Proto-humans certainly had a capacity to measure resemblance between items, since it is central to many other capacities, like "memory, categorization, decision making, problem solving, and reasoning": it is a "factotum to cognition."[13] Since this capacity existed and served as a go-to cognitive tool for many tasks, its recruitment to represent direct evidence would have been unsurprising.

C. The ERM as a Cartesian Device implementation

Suppose I use an ERM. As my visual module outputs THERE IS RED, my ERM-powered metacognitive module outputs THERE IS *RED* IN ME, and estimates that *red* gives me direct evidence for what it distinctively resembles.

Here is a plausible assumption about resemblance estimates: we use representational similarity as a proxy for item similarity. It is plausible, because it

[12] I previously used "maximal resemblance" (Kammerer 2019c; 2021), but "distinctive resemblance" is clearer.

[13] Larkey and Markman (2005, 1061).

seems assumed by leading theories of resemblance judgments: "most writers" assume that judgments of "similarity between objects come from some sort of comparison between their representations."[14] Representational similarity is used as a proxy for item similarity according to transformational accounts of resemblance judgments, such as the representational distortion accounts,[15] but it also plays a role according to competing structure-mapping models.

Using representational similarity as a proxy for item similarity, the ERM determines that *red* gives an agent direct evidence for the presence of red: since *RED* recruits RED's vehicle, both representations are highly similar, and of all representations formed with * . . . *, *RED* is the most similar to *RED*. This generalizes to other outputs.

The ERM makes two other crucial conclusions. First, it concludes that *red* gives the agent revealing direct evidence for the presence of *red* itself: it grasps *red* as self-revealing.

Indeed, *red* is trivially as similar *to itself* as possible. The same applies, at the representational level, to *RED*: identity implies distinctive resemblance. So, the ERM considers that *red* distinctively resembles itself, and therefore gives direct evidence for itself. This generalizes to other states *x*: the ERM represents direct evidence as self-presenting. Now,*RED*, like other concepts *x*, characterizes its referent as primitive: as having a simple, uncomposed nature captured in this representation. When the ERM represents *red* as self-presenting, it represents it as giving direct evidence for its own presence, grasped with *RED*, and it understands *RED* as capturing the nature of this direct evidence. Therefore, it represents *red* as self-revealing. Again, the point generalizes to other states *x*: the ERM represents direct evidence as self-revealing.

The ERM's second crucial conclusion is that, if an agent has direct evidence for the having of *red*, the evidence cannot be deceptive: it grasps *red* as infallibly directly presented.

Indeed, what is needed, according to the ERM, to have direct evidence for the having of *red*? To be in a state *x* distinctively resembling *red*. Using representational similarity to assess item similarity, the ERM concludes that the having of *red* itself is what is required to have direct evidence for the having of *red*: of all states *x*, *red* resembles *red* the most, since *RED*,

[14] Blough (2001).

[15] On these accounts, "the similarity between a pair of items is inversely related to the number of basic transformations involved in distorting the representation of one item into the representation of the other item" (Larkey and Markman 2005, 1062).

trivially, is the concept formed with *…* that resembles *RED* the most. Any concept *x* which is not *RED* resembles *RED* less than *RED* does. This generalizes to other states *x*: the ERM represents direct evidence as infallibly directly presented.

Let us summarize. Using an ERM to represent direct evidence, we represent direct evidence as primitive. We represent it as providing evidence for things beyond itself (e.g., we represent *red* as giving evidence for red). We represent it as self-revealing and as infallibly directly presented. So, using an ERM, we implement a Cartesian Device.

The ERM is a reasonably simple mechanism. It recruits components that plausibly preexisted in proto-humans, like sensory concepts and a measure of resemblance. This gives preliminary plausibility to *Implementation*, the second tenet of Meta-Cartesianism:

> *Implementation*: We implement a Cartesian Device with an Evidence-by-Resemblance Mechanism.

If both *Implementation* and *Architecture* are true, concepts *x* manipulated by the ERM, like *RED*, are our concepts of direct evidence *and* our phenomenal concepts. The operator *… * is a *phenomenal operator*, taking non-phenomenal concepts, like sensory concepts (e.g., RED), and transforming them in phenomenal concepts representing phenomenal properties (e.g., *RED*). For now, and in what immediately follows, I discuss only *sensory* phenomenal concepts, formed by applying the phenomenal operator to sensory concepts. I consider non-sensory phenomenal concepts at the end of the next chapter.

D. Resemblance views of perception

Implementation states that we implement a Cartesian Device through an ERM. One thing makes the claim plausible: the fact that it makes a specific confirmed prediction. Namely, it predicts that mental states represented as providing direct evidence, like sensory perceptions, will be intuitively viewed as resembling what they give evidence for—the objects of sensory perceptions. This is exactly what happens: the idea that sensory perceptions resemble their objects was historically extremely popular, and many who rejected it conceded that it is highly intuitive.

Aristotle held a resemblance view of perception: he thought that perceived objects impact our sensory faculties and make them become *like* them.[16] This was the received view of perception in Arabo-Islamic medieval philosophy and Western medieval philosophy, at least until the 14th century.[17] Prominent non-Aristotelian Ancient philosophers had similar ideas. Empedocles and Democritus, followed by Epicurus and his school, held that perception relied on transmitting "*simulacras*" or "replicas" of the perceived object to the perceiving mind. The sophist Gorgias thought that, in vision, the soul was "impressed" and "moulded" by the thing seen—that vision "inscribes on the mind likeness of the things seen." Similar ideas are also frequently found in the works of Plato and the Stoics.[18]

We find comparable ideas elsewhere. In the Indian Buddhist tradition, the Sautrāntika school is associated with the view that perceptual awareness is related to the perceived object by causation and *resemblance* (sārūpya).[19] Highly influential members of the school held the view. Dignāga (fifth–sixth century CE), among the most important Buddhist thinkers, thought that "perceptual awareness" has "likeness" to the external perceived object.[20] His master, Vasubandhu (fourth–fifth century CE), also thought that perceptual awareness resembles its object.[21] For Birgit Kellner, an expert in Buddhist philosophy, the resemblance of view of perception was probably "not exclusive to Sautrāntikas," and "its close association with Sautrāntika emerged as a result of its prominent place in [. . .] Buddhist epistemology."[22]

Many, in the West and in the Indian tradition, rejected the resemblance view of perception, but conceded that it is intuitive. Dharmakīrti (sixth or seventh century CE), from the Yogācāra school, defended idealism. He thought that external objects, being non-existent, could not resemble perceptions. However, his argument for idealism used the resemblance theory of perception as a background assumption: he thought that perception, possessing only its own form (i.e., resembling only itself), is merely, and always, directly aware of itself (i.e., in my vocabulary, only provides direct evidence for itself).[23]

[16] See the *De Anima* ii 5, 418a3–6 (for a discussion, see Shields 2020a; and more precisely 2020b).
[17] King (2007a).
[18] References and quotes in this paragraph are all from Caston (2015, 33–34).
[19] Kellner (2015, 313).
[20] Kellner (2014, 276).
[21] Kellner (2014, 279–81).
[22] Kellner (2015, 313).
[23] Kellner (2015, 315).

Since Galileo and Descartes, Western philosophers have often rejected the resemblance view of perception, at least for sensory perceptions of qualities like color or warmth. Still, this Early Modern stance came hand-in-hand with the view that perceived objects *differ* from what they perceptually appear to be. Galileo, Descartes, Hobbes, Boyle, and Locke diverged on the exact status of secondary qualities, but they all thought that "secondary qualities like colour and sound are *not* in physical objects in the way they sensorily appear to be."[24] This is in agreement with the view that, intuitively, perceptions resemble what they present: objects' actual qualities do not really resemble perceptions, but they are not as they perceptually appear.

Descartes, in 1644, in his *Principles of Philosophy*, made it very clear: he denied that perceptions resemble perceived objects, while granting that the resemblance view is universally intuitive:

> all of us have, from our early childhood, judged that all the objects of our sense-perception are things existing outside our minds and closely resembling our sensation.[25]

So, Early Modern Western philosophers denied the resemblance of perception to actual external objects, while often acknowledging that perceptions make objects resembling them appear to us, and denying that *real* external objects really resemble perception and that they are as they perceptually appear to be.

Therefore, *Implementation* correctly predicts that the resemblance view of perception is highly intuitive, which gives plausibility to the thesis. At the same time, it does not entail that this resemblance view is correct: in fact, Meta-Cartesianism implies that it is false.

5.5. Meta-Cartesian illusionism

Architecture and *Implementation* both have preliminary plausibility. If they are true, we use an ERM-powered Cartesian Device in evidential and phenomenal cognition: our phenomenal concepts are concepts of direct evidence. However, taken together, these theses remain neutral on whether

[24] Simmons (2015, 84).
[25] Descartes (1985a, 216).

phenomenal concepts are satisfied: they are neutral between illusionism and realism. To get to Meta-Cartesianism, we must add the third tenet of the view:

Illusionism: There are no phenomenal experiences: our Cartesian Device's representations of phenomenal experiences are misrepresentations

I will not argue again for illusionism here, as this has already been done. I will rather focus on the right way to understand the illusionist dimension of Meta-Cartesianism.

For Meta-Cartesianism, many things are real. Inferential and non-inferential belief-formation processes are real. The second are really tracked by our ERM-implemented Cartesian Device, which really deploys phenomenal concepts characterizing their referents as primitive, self-revealing, and infallibly directly presented direct evidence. However, this characterization is a mischaracterization. There is no primitive, self-revealing, and infallibly directly presented direct evidence. Phenomenal concepts mischaracterize what they track: our real non-inferential belief-formation processes, such as broadcast sensory outputs. These processes are not primitive. They do not resemble the qualities represented by our sensory concepts, and if they provide evidence about these qualities, it is *not* by resembling them. We do not have, with these processes, the unique epistemic relation our phenomenal concepts depict. These processes are not self-revealing: they can happen in us while we lack revealing evidence that they do. They are *not* infallibly directly presented: we can have deceptive evidence, independent of beliefs, that they happen in us.

One could grant that the conjunction of *Architecture* and *Implementation*—the view that we use an ERM-powered Cartesian Device for evidential and phenomenal cognition—is plausible, but maintain that *Illusionism*—and therefore, Meta-Cartesianism itself—is not. For instance, Meta-Cartesianism implies that the phenomenal concepts deployed by our ERM-powered Cartesian Device are mischaracterizations, and one could deny that it could ever have been adaptive to develop such mischaracterizing representations.

Some, including some illusionists, seem to think that it is only plausible that we evolutionarily acquired mischaracterizing representations if those are adaptive and fitness-increasing *as mischaracterizations*.[26] I disagree.

[26] Humphrey (2011).

Meta-Cartesianism is evolutionarily plausible if the possession of the right representations, which we have independent reasons to think are illusory, is adaptive, but there is no need for these representations to be adaptive *because* they are illusory. At most, their illusory nature should not be *sufficiently maladaptive* to cancel out their overall adaptive nature. That it is the case here is plausible, since the mischaracterizing representations posited by Meta-Cartesianism still provide a convenient rough grasp of belief-formation processes under an evidential guise; a grasp which is probably good enough in most contexts, and only generates problems in specific theoretical situations that were inexistent or irrelevant in our ancestral environment.

At this stage, Meta-Cartesianism's plausibility remains modest. Three things must be kept in mind.

First, most of the view's plausibility comes from the fact that it solves the illusion problem. This is the focus of the next chapter.

Second, Meta-Cartesianism is not an exhaustive view of phenomenal and evidential cognition. It states that phenomenal and evidential cognition use an ERM-powered Cartesian Device, but it does not imply that they do not also mobilize other mechanisms or devices. These forms of cognition could be messy, plural, and highly fractionated[27]: Meta-Cartesianism could still be true, as long as an ERM-powered Cartesian Device plays a sufficiently central role in them. I also did not speculate on the links between evidential cognition, which represents reasons to believe, and other forms of normative cognition, like those representing reasons to act. I find it highly plausible that those are tightly related, but this is independent of Meta-Cartesianism.

Third, Meta-Cartesianism is speculative, but it helps to separate the view from its underlying key insight. This insight is at the basis of the evidential approach: it is the idea that the cognitive apparatus we use to think about *evidence*—what is presented to us, what we have reasons to believe, etc.—is tightly linked with the cognitive apparatus we use to think about phenomenal experiences, and that this explains why we grasp phenomenal experiences as entities standing in a particularly intimate evidential relation with us. Meta-Cartesianism is one specific way of fleshing out this key insight. It might turn out to be incorrect, but I am deeply convinced that the right view of phenomenal cognition will have to do justice to this insight. I would welcome the development of rival views building on it.

[27] Schwitzgebel (2012).

5.6. Contrasting Meta-Cartesianism

Meta-Cartesianism upholds the *Common Core* shared by Graziano's AST and Pereboom's QIH: in virtue of hard-wired and innate features of our introspection, we represent some of our mental states as having special properties which they lack. Of course, Meta-Cartesianism goes beyond this *Common Core*.

In some respects, Meta-Cartesianism is close to Graziano's AST. For AST, we represent phenomenal consciousness by modeling what we attend to. For Meta-Cartesianism, we represent phenomenal consciousness by modeling what directly appears to us: our reliable non-inferential belief-formation processes. *Contra* AST, Meta-Cartesianism is a rich-illusion view. It does not view our representations of phenomenal experiences as simply schematic and incomplete, but as *positive* misrepresentations.

This makes Meta-Cartesianism closer to Pereboom's views. From Pereboom's PIH, Meta-Cartesianism borrows the idea that we represent phenomenal properties as primitive. However, it goes beyond it: we do not just grasp them as primitive properties, but as primitive properties *with unique evidential features.*

Meta-Cartesianism has things in common with other conceptions. It views our phenomenal concepts as redeployments of modular concepts (such as sensory concepts) with a phenomenal operator. The idea that phenomenal concepts result from this sort of recombination is widespread among realists[28] as well as illusionists.[29] It aligns well with the popular view that introspection of sensory perceptual states consists of the deployment of sensory capacities, with a conceptual twist.[30]

Contrary to views appealing to introspective opacity, Meta-Cartesianism states that what makes our introspection of phenomenal experiences peculiar is not only that it is limited in terms of what it can access: it is that it positively mischaracterizes, in evidential terms, some of our real mental states.

Contrary to PRM, which aims at explaining phenomenal experience, but not mystery intuitions, Meta-Cartesianism aims exclusively at explaining

[28] Papineau (2002); Carruthers (2005b); Aydede and Güzeldere (2005).
[29] Frankish (2016, 36).
[30] Wu (2023).

the appearance of phenomenal experiences and mystery intuitions. Its key explanatory posit is not just the monitoring of our evidentially relevant internal states and their characterization as evidential, but the substantive representation of certain states as peculiar direct evidential states, characterized as primitive, self-revealing, and infallibly directly presented.

Contrary to designed certainty views of Schwarz or A. Clark et al., Meta-Cartesianism does not cash out the peculiarity of phenomenal introspection in terms of the certainty of its output. For Meta-Cartesianism, the peculiarity of phenomenal introspection is primarily a *represented* peculiarity. Phenomenal introspection is not peculiar because its outputs are *psychologically* certain: Meta-Cartesianism does not imply that these outputs are psychologically certain. In fact, it rather seems plausible that the psychological confidence attached to introspective outputs varies. Introspective outputs concerning central features of sensory phenomenal experiences, corresponding to easily tracked salient non-inferential belief-formation processes, are probably very confident—e.g., we are extremely confident when we introspect a strong phenomenal pain or a salient experience of a bright red spot. In contrast, introspective outputs concerning less central features of experiences, or marginal experiences (see the next chapter on cognitive experiences), seem less confident, notably as they correspond to harder-to-track non-inferential processes, eliciting less decisive introspective outputs.

So, contrary to designed certainty views, Meta-Cartesianism does not view the outputs of phenomenal introspection as psychologically certain. What happens instead is this: phenomenal cognition represents phenomenal experiences as standing in a unique epistemic relation to subjects. To cash it out in terms of certainty: it represents phenomenal experiences as standing in a relation to subjects that affords *epistemological* certainty.

As noted earlier, Meta-Cartesianism is the latest iteration of ideas I developed over the last decade. Some ancestors of Meta-Cartesianism stated that something close to what I call here a Cartesian Device was implemented by a modular naïve theory of mind and naïve epistemology, holding certain (false, but convenient) naïve principles regarding phenomenal experiences and their epistemic relations to subjects.[31] This is technically compatible with Meta-Cartesianism, if we understand the relevant naïve theories as modular

[31] See notably Kammerer (2016b; 2019a; 2021). See also Kammerer (2016c), although this early version only accounted for the fact that phenomenal experiences are represented as infallibly directly presented, not as self-revealing. The idea of the ERM was introduced a bit later (Kammerer 2019c).

and implicit. Still, "theory" connotes something explicit and controlled, which is why I now prefer avoiding these formulations.

* * *

I presented Meta-Cartesianism and made a preliminary case for the view. In the next chapter, I explain how it solves the illusion problem.

6

Leaving the House of Mirrors

Solving the Illusion Problem

In this chapter, I argue that Meta-Cartesianism solves the illusion problem. It explains our intuition of obviousness and solves the illusion meta-problem. It explains our intuition of undetectability. It explains why our intuition of irreducibility about the phenomenal keeps bothering us. It accounts for miscellaneous intuitions and for the appearance of non-sensory phenomenality. It explains our intuitions of significance. Objections to Meta-Cartesianism can be answered.

6.1. The intuition of obviousness and the illusion meta-problem

I take some things for granted. Meta-Cartesianism upholds the *Common Core*. As a result, it explains our shared and compelling introspective disposition to judge that we enter certain mental states, endowed with certain properties. As it views our representations of phenomenal experiences as representations of *primitive* direct evidence, it explains why phenomenal consciousness is conceptually distinct from access-consciousness or self-consciousness. Like Pereboom's *Primitive Inaccuracy Hypothesis* (PIH), it explains why we have an intuition of irreducibility about the phenomenal. Of course, we still need to explain why this intuition keeps bothering us.

Having assumed this, I must now show that Meta-Cartesianism succeeds where other views fail: it explains our intuition of obviousness and solves the illusion meta-problem.

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0007

A. Explaining the intuition of obviousness

Existing views do not explain our shared, compelling, and conceptual intuition of obviousness. This includes the intuition that phenomenal experiences are *introspectively self-revealing*: if a subject has a phenomenal experience, its essence introspectively appears to the subject. It also includes the intuition that they are *infallibly introspectively appearing*: if it introspectively appears to a subject that they have an experience with a phenomenal property, they have an experience with this property.

Meta-Cartesianism is tailored to do it. It posits phenomenal concepts representing phenomenal experiences as primitive, self-revealing, and infallibly directly presented direct evidence. If we assume, as is highly plausible, that "introspective appearance" is analyzable as "a direct justification that P, where P is about the subject's own present mental state," Meta-Cartesianism explains our conceptual intuition that phenomenal experiences are *introspectively self-revealing* and *infallibly introspectively appearing*. Moreover, for Meta-Cartesianism, our phenomenal concepts are governed by a shared innate and inflexible module. This explains that our intuition of obviousness is shared and compelling.

Meta-Cartesianism explains our intuition of obviousness. This is unsurprising: the link it posits between evidential and phenomenal cognition is designed to explain our grasp of the phenomenal as obvious.

B. Obviousness and kinds of appearances

Where does this explanation leave us? Return to the three typess of appearance distinguished in Chapter 1: psychological appearances (dispositions to believe), epistemic appearances (reasons to believe), and phenomenal appearances (epistemic appearances given through a phenomenal medium).

We have an intuition of obviousness: phenomenal experiences *psychologically* appear introspectively obvious to us. To explain this, Meta-Cartesianism appeals to features of our evidential cognition and our innate concept of direct evidence. This last concept is a concept of a certain kind of *epistemic* appearance. So, Meta-Cartesianism explains the *psychological* appearance of phenomenal consciousness as *obvious* by our innate grasp of specific *epistemic* appearances. The explanation works because the intuitive

obviousness of the phenomenal can be cashed out in terms of specific epistemic appearances: on a plausible analysis, a mental state is introspectively self-revealing when its essence always directly epistemically appears to the subject in it. It is infallibly introspectively appearing when the direct epistemic appearances the subject has about it are infallible.

For Meta-Cartesianism, phenomenal concepts represent phenomenal experiences *as* direct evidence—a very specific sort of epistemic appearances. More precisely, they represent them as *phenomenal appearances*: epistemic appearances given in a phenomenal medium.

Meta-Cartesianism correctly predicts that only direct evidence about our own phenomenal experiences is represented as always trustworthy: we only have an intuition of obviousness about *introspective* appearances of phenomenal experiences. There is no obstacle to representing deceptive non-introspective *indirect* evidence about our phenomenal experiences: we easily think of a brain-imagery device feeding us misleading information about our own current experience. There is also no obstacle to representing misleading *direct evidence* about someone else's phenomenal experiences: intuitively, a skilled actor can fool me and make me wrongly 'feel' that they are in pain, thus providing me with deceptive direct evidence about their experiences. Finally, there is no obstacle to representing misleading direct evidence about our own past or future experiences: intuitively, I can wrongly feel that I will experience deep sadness tomorrow, even if in fact I experience nothing of the kind when the day comes.

C. Phenomenal introspection as-it-is and as-represented

We find ourselves in a house of mirrors. To better understand it, distinguish between two sorts of phenomenal introspection. There is *phenomenal-introspection-as-it-is*: the real process, involving our ERM-powered Cartesian Device, by which we represent our phenomenal experiences. It is a modular process, among others. It involves no 'magic'. It is only peculiar to the extent that it mischaracterizes what it tracks as having unique evidential properties.

These unique evidential properties do not exist. There is no uniquely intimate epistemic relation between subjects and their phenomenal experiences. Those are merely represented. As a result, aside from phenomenal-introspection-as-it-is, there is *phenomenal-introspection-as-represented*. It involves this uniquely intimate epistemic relation between us and our

phenomenal experiences. This is where the 'magic' lies—a non-existent, merely represented sort of magic.

Phenomenal-introspection-as-it-is has no mysterious peculiarities; phenomenal-introspection-as-represented is otherworldly. Within phenomenal-introspection-as-it-is, phenomenal experiences are neither self-revealing nor infallibly introspectively appearing; within phenomenal-introspection-as-represented, they are.

In phenomenal-introspection-as-it-is, the *target* of introspection—non-inferential belief-formation processes—differs from its medium—phenomenal concepts. Phenomenal concepts target and mischaracterize these processes. In phenomenal-introspection-as-represented, phenomenal experiences are both the target (what appears to the subject) and the medium (what provides the introspective appearance to the subject). Intuitively, we cannot isolate the way phenomenal experiences *appear to us introspectively*, and distinguish it from the way they *are*. There seems to be no specific *introspective phenomenology* distinct from *sensory phenomenology*.[1]

D. Introspective appearances and introspective beliefs

Intuitively, introspective *appearances* about phenomenal experiences are infallible, but introspective *beliefs* are not.[2] For instance, we can easily admit that, in the fraternity case, the blindfolded student briefly formed a false introspective belief, mistaking a feeling of cold for a feeling of pain.

Meta-Cartesianism explains this: since what is intuitively impossible are only misleading introspective appearances about our own phenomenal experiences, false *beliefs* about these experiences are intuitively admissible *if* they depart from what introspectively appears to us. It seems to us that we can make what philosophers have called "errors of judgments,"[3] or "basing errors,"[4]about our own phenomenal experiences: we can form false beliefs about our present phenomenal experiences if, focusing on them, we are biased, inattentive, or fail to master the concepts we deploy, etc. We then fail to base our judgments on introspective appearances and end up with false beliefs. For the same reason, what has been called "errors of ignorance" or

[1] Shoemaker (1994); Lycan (1996); Siewert (2012).
[2] Pace other views criticized in Chapter 4 (A. Clark et al. 2019; Schwarz 2018; Shabasson 2022).
[3] Hill (1991, 127–28).
[4] Smithies (2013, 1180).

"brute errors" are intuitively impossible: introspective error cannot stem from misleading introspective appearances. By contrast, in perception, all sorts of errors seem possible.

Intuitively, the blindfolded student made a mistake, but only because he expected pain. His judgment thus failed to correspond to how things introspectively appeared to him. However, we have the impression that, if he had properly paid attention to what happened to him, and if he had mastered his fear and had not expected pain, he would have correctly judged that he was feeling cold, not pain.

So, Meta-Cartesianism does not imply that introspective beliefs are infallible. It does not imply that they are intuited as infallible. As noted previously when discussing designed certainty views, it does not imply that introspective outputs or introspective beliefs are psychologically certain. It is compatible with the plausible idea that different introspective beliefs are held with various degrees of confidence, because our metacognitive module forms more confident introspective outputs in some cases than in others. Plausibly, outputs about central, coarse-grained, and salient phenomenal features of intense and familiar experiences are held with great confidence; outputs about marginal and subtle phenomenal features of unfamiliar and feeble experiences are not.

So, intuitively, we view introspective appearances as self-revealing and infallible, but think that introspective beliefs are fallible, and hold them with various degrees of confidence. This generates an apparent paradox. Why do we cling to the idea that introspective appearances are self-revealing and infallible, since we see that introspective beliefs can be erroneous, incomplete, and hesitant? Meta-Cartesianism explains this apparent paradox. We *essentially* represent phenomenal experiences as introspectively obvious. Facing our own introspective mistakes and hesitations, we are conceptually forced, so to speak, to maintain that introspective appearances are self-revealing and infallible, and we are led to systematically put the blame on factors other than introspective appearances—bias, lack of attention, irrationality, insufficient conceptual mastery, etc.

E. Solving the illusion meta-problem

Trying to make sense of compelling illusions, we deploy evidential cognition. Using our Cartesian Device, we represent illusions as situations in which we have deceptive direct evidence. For instance, suppose we encounter the Müller-Lyer illusion, knowing it is an illusion. To make sense of what happens, we represent that we have deceptive direct evidence that the two lines have different lengths: we have a phenomenal experience of two lines with different lengths, even if the lines have similar lengths.

For Meta-Cartesianism, phenomenal and evidential cognition are intertwined. This makes our intuitive way of dealing with illusions inapplicable to the illusion of phenomenality: as we represent direct evidence about our own phenomenal experiences as infallible, illusionism seems incoherent. Here, however, we must be careful. The illusionist view is that phenomenal experiences do not exist, but appear to exist. Since there are three concepts of appearances (psychological, epistemic, phenomenal), the view can be read in three ways.

First, if "appear" means "phenomenally appear," illusionism states that there are no phenomenal experiences, although it phenomenally appears to us that there are some. This is blatantly incoherent. We do not need Meta-Cartesianism to see why. No one defends this view.

Second, if "appear" means "epistemically appear," illusionism states that there are no phenomenal experiences, although we have some reasons to believe there are some. This view is coherent and true, even if it is not my primary definition of illusionism. Still, given our evidential cognition, this view generates deep tensions. Suppose you admit that phenomenal consciousness epistemically appears to exist without existing. Is the deceptive evidence we have for its existence direct or indirect? Saying that it is only indirect, and depends entirely on our other beliefs, seems implausible. This is roughly what theoretical eliminativists are claiming, and I think that this is unconvincing: introspection gives us reasons to believe that we currently have phenomenal experiences, and these reasons do not depend on our beliefs. So, when we consider the view that the phenomenal epistemically appears to exist without existing, we think that the relevant epistemic appearance must be direct: we must have deceptive direct evidence for the existence of our current phenomenal experiences. However, given how our evidential cognition works, we represent direct evidence for our own current phenomenal

experiences as always trustworthy. So, in this epistemic reading, illusionism seems incoherent.

Third, if "appear" means "psychologically appear," illusionism states that phenomenal experiences do not exist, although we tend to believe they do. This is how I defined illusionism. This view is coherent and true. Still, a tension quickly arises when we consider it further. We wonder whether the psychological appearance of phenomenal consciousness is accompanied by a corresponding epistemic appearance. If not, our tendency to believe that we have phenomenal experiences is baseless and irrational. As everyone shares the tendency, and almost everyone the belief, this means that we are all massively irrational and virtually insane. It seems hard to avoid conceding that this psychological appearance comes with an epistemic appearance. This leads to the second reading of illusionism, which, as I just showed, seems incoherent.

This is the house of mirrors we are trying to escape. The exit path is narrow and hard to find. Still, it exists. We must close our eyes and follow the handrail. Denying phenomenal experiences, we must suspend our trust in our innate, automatic, and intuitive understanding of evidence. Direct evidence does not need to be grounded in phenomenal experiences. Maybe a mere disposition to believe something is sufficient for direct evidence? Maybe some other mental state does the job? (On this, see Chapter 10.) At any rate, once we admit that direct evidence does not need to be grounded in phenomenal experiences—thus going against compelling evidential intuitions—we appreciate the coherence of the claim that phenomenal experiences appear to exist, psychologically and epistemically, without existing.

Illusionism is coherent, but it takes effort to appreciate its coherence, since we can only do it, and potentially accept the view, if we appropriately disregard two interconnected processes: phenomenal introspection and our intuitive evidential cognition. This requires going against deep tendencies. It is the only way to leave the house of mirrors. Having left, we understand why it was so hard to leave, why many people find illusionism incoherent and insane, and why the illusion of phenomenal consciousness is exceptional and uniquely strong. We also understand why it will always be so extraordinarily tempting to go back inside and be lured again by the seductive reflections of the phenomenal mind.

6.2. Explaining the intuition of undetectability

Meta-Cartesianism also explains our shared, compelling, and conceptual intuition of undetectability—the intuition that phenomenal consciousness is not directly detectable outside of introspection. The explanation has three components.

A. Acquiring phenomenal concepts

How are phenomenal concepts acquired? For Meta-Cartesianism, our ERM forms them by redeploying non-phenomenal concepts, like sensory concepts (e.g., RED), with the phenomenal operator. This redeployment produces phenomenal concepts (e.g., *RED*): concepts of primitive direct evidence for whatever the initial non-phenomenal concept represents.

Focus on *sensory* phenomenal concepts. Being formed by redeploying sensory concepts, their formation requires the antecedent possession of the right sensory concepts. However, sensory concepts normally depend on the activity of corresponding sensory capacities. Frank Jackson's Mary, in her black-and-white room, does not possess the sensory concept RED, and thus cannot form *RED*. Humans, lacking the echolocative capacities of bats, do not have echolocative sensory concepts and thus cannot form phenomenal concepts of echolocative sensations.

This partially explains the undetectability intuition: in normal cases, we cannot form concepts of phenomenal experiences corresponding to sensory capacities we lack. These experiences seem beyond our comprehension.

B. Mindreading as inference

Still, this does not fully explain our intuition of undetectability, which also concerns familiar experiences of others. Suppose I am sitting in a car with Tobias. We both look at the road. I think his vision works just like mine. Still, his experiences do not seem directly detectable to me, contrary to mine. I only know about them by observing his behavior, gaze, etc., and *inferring* what he is visually experiencing.

There is a natural explanation for this. It does not follow from the tenets of Meta-Cartesianism, but it is compatible with them. Plausibly, as a matter

of fact, our ERM-powered Cartesian Device's introspective deployment is highly automatic and encapsulated, but its deployment in mindreading is not as automatic and encapsulated. It is relatively more sensitive to central beliefs. If so, phenomenal experiences will seem less directly detectable in mindreading than in introspection. The introspective output I EXPERIENCE RED will be reached non-inferentially, while the mindreading output TOBIAS EXPERIENCES RED will typically require some personal-level inferences on the basis of observed behavior.

C. Relative undetectability

Still, even this cannot be the whole story. Consider heartfelt empathy. My friend Paul screams while holding his foot. As a result, I understand, immediately, that he experiences pain—I feel it, without any inference. However, even then, my grasp of his pain seems indirect compared to my grasp of my own pain when I have it. If Paul were a skilled actor, he could fool me.

To understand why this happens, come back to the intuition of obviousness. We represent phenomenal experiences as standing in a uniquely intimate epistemic relation to the subject who has them. This relation can be contrasted to the relation a subject has with the experiences of others, or with their own past and future experiences.

Therefore, even when I experience empathy for Paul, his pain just does not seem introspectively self-revealed nor infallibly introspectively presented to me, even if my own *empathetic experience* seems self-revealed and infallibly introspectively presented to me, and even if Paul's pain experience seems self-revealed and infallibly introspectively presented to *him*. In empathy, I represent Paul's pain non-inferentially, but I still grasp my epistemic relation to his pain as mediated, precisely, by my empathetic experience, a potentially deceptive medium. Trying to imagine a perfectly immediate epistemic relation to Paul's pain, I end up imagining that his pain literally becomes *mine*—if this is imaginable.[5]

This completes the explanation of the intuition of undetectability. Phenomenal experiences seem *relatively* undetectable outside of introspection because, outside of introspection, we represent our epistemic relation to

[5] For the idea that two subjects can share the same token experience, using the case of craniopagus twins, see Cochrane (2021).

them as mediated by inferences or empathetic experiences. By contrast, we represent our introspective relation to our experiences as immediate. This explains a sense of *relative* undetectability.

Intuitively, many non-phenomenal properties are also accessed mediately: e.g., we only access external sensory qualities mediately, through sensation. However, the phenomenal seems special because, while we think that our relation to external sensory properties is mediated and fallible, this relation seems as good as it gets: for us, there is no more direct and secure way to access these properties. In the case of phenomenal experiences, however, we think that a more direct and immediate access is always available to the experiencing subject. As a result, we tend to view phenomenal experiences as part of a closed-off subjective world to which even empathy offers only second-rate access.

6.3. Explaining the intuition of irreducibility

For Meta-Cartesianism, as for Pereboom's PIH, we represent phenomenal properties as primitive. This explains why a first intuition of irreducibility about the phenomenal arises. However, we still need to explain why this intuition keeps bothering us.

For instance, why does it bother us much more than our intuition of irreducibility about primitive colors? We easily discard irreducibility intuitions about primitive colors as *irrelevant*, and concerning only non-existent primitive qualities. It is "one of the clichés of the history of philosophy" that we have been doing this since the 17th century.[6] According to the "relocation story,"[7] we used to believe, with common sense and Aristotle, in a world full of primitive sensory qualities. Galileo and Descartes convinced us that the physical world lacks primitive sensory qualities and contains only non-qualitative and mathematically describable properties. We came to see primitive qualities as only present in our perceptions: using a "displacement maneuver,"[8] we *relocated* these qualities. We still discuss the metaphysics of sensory qualities, but the mainstream view is that primitive sensory qualities

[6] Shoemaker (1990, 110).
[7] Rosenthal (1999).
[8] Shoemaker (1990, 111).

do not really exist in the outside world, and irreducibility intuitions about them do not bother us.

Meta-Cartesianism explains why, in contrast, phenomenal irreducibility intuitions keep bothering us. It happens, precisely, because we cannot intuitively do about primitive phenomenal properties what we did about primitive sensory qualities. We struggle to admit that they do not exist but seem to exist, because we view them as introspectively obvious.

Indeed, suppose we envision the view that phenomenal properties do not exist, or at least not as they appear to be. We try to make sense of this view, but it runs counter to deep tendencies we have. To make sense of the view without supposing that we are all insane, we must admit that our own phenomenal properties *psychologically* and *epistemically* appear to exist—in a non-inferential and direct manner. So, we must accept that we have deceptive direct evidence about them. However, our Cartesian Device's functioning leads us to conclude that this is impossible. Introspective direct evidence about phenomenal properties is always trustworthy and it discloses the essence of these properties. Primitive phenomenal properties must exist as they appear. Therefore, while irreducibility intuitions arise for every property represented as primitive, it is only for phenomenal properties that we cannot discard them as irrelevant, because of our intuition of introspective obviousness about the phenomenal.

This explains why, while many think there is a *hard problem of consciousness*, few think there is a similar *hard problem of sensory qualities*—say, a hard problem of colors. Interestingly, it also explains why a hard problem of sensory qualities emerges for some. Those who believe that, in perception, things are *as they seem*—naïve realists—will apprehend sensory qualities as primitive, as they appear to be. They will face a hard problem—an "objective explanatory gap"—comparable to the hard problem of consciousness.[9]

[9] Shoemaker (2003, 254); Fish (2008, n. 7).
[10] Levine (2001, 6–9); Kriegel (2005).

6.4. Explaining miscellaneous intuitions

A. Phenomenal experiences as qualitative and subjective

Meta-Cartesianism explains other intuitions. First, it explains why phenomenal experiences appear to have two aspects: a *qualitative* aspect, about which we have an intuition of ineffability, and a *subjective* aspect—a *for-me-ness*.[10]

Take phenomenal concepts *RED* and *GREEN*. They grasp their referents as making different sensory qualities, represented with different sensory concepts, appear to the subject. The aspect under which they are represented as distinct is also grasped as distinctively resembling distinct sensory qualities. It is this aspect we call "qualitative."

Sensory concepts grasp sensory qualities as primitive qualities, implicitly placing them in a quality space, corresponding to the discriminations and classifications our sensory capacities enable. This causes an intuition of *ineffability* about external sensory qualities, since we know much about them that we cannot articulate linguistically. However, this ineffability intuition can be discarded as ultimately irrelevant once we admit that external sensory qualities do not exist as they seem to. We can conclude that, whatever ordinary physical color-related properties red and green things have, they might be entirely expressible in language (at least in principle) even if *we* cannot easily express everything we know about them. As we also place phenomenal qualities in a comparable quality space, we have the same ineffability intuition about them. Still, as previously with the irreducibility intuition, the problem is that we cannot discard it as irrelevant and bearing merely on nonexistent properties.

Finally, we grasp *red* and *green* as *qualitatively* different, but also as having something in common: indeed, both are represented with phenomenal concepts formed with the same phenomenal operator. We grasp both as states presenting something *to* the subject and standing in a uniquely intimate epistemic relation *with* the subject. It is this aspect of phenomenal experiences we call its "subjective" dimension—its "for-me-ness."

[11] Dretske (2003, 73).
[12] Dretske (1995); Tye (1995); Siewert (1998); Chalmers (2004).

B. Phenomenal experiences as representational, meta-representational, and self-representational

Another intuition, which I did not discuss so far, is that many phenomenal experiences, in virtue of their phenomenal character, *represent* worldly features. Relatedly, we classify phenomenal experiences in terms of the worldly features they represent: we talk of an experience of blue, an olfactory experience of coffee, an auditory experience of a middle C, etc.[11] This intuition and this classification have been used to support *representationalism* about phenomenal consciousness.[12]

Meta-Cartesianism, being an illusionist view, rejects representationalism about the phenomenal. Still, it explains the intuitive pull of the view: given Meta-Cartesianism, we essentially grasp phenomenal experiences as epistemically presenting certain features to us, and we individuate different phenomenal experiences in terms of the features they thus present.

Many also intuit that phenomenal experiences are states we are aware *of*. This intuition typically motivates *higher-order representationalism*.[13] Others insist that, intuitively, what makes it so that we are aware of our phenomenal experiences is the way phenomenal experiences are in themselves: phenomenal experiences, in virtue of what they are, represent themselves. This motivates *self-representationalism*.[14]

Again, since Meta-Cartesianism is an illusionist view, it rejects these views. Still, it explains the intuitions in their favor. Our phenomenal concepts represent phenomenal experiences as epistemically presenting certain features, but also as always being themselves presented to us—hence, higher-order representationalist intuitions. They represent phenomenal experiences as being presented to us in virtue of what they are in themselves, since we grasp phenomenal experiences as self-revealing—hence, self-representationalist intuitions.

6.5. Phenomenality beyond the sensory

I showed that Meta-Cartesianism explains the illusion of *sensory* phenom-

[13] Armstrong (1980); Lycan (1996); Rosenthal (2005).
[14] Kriegel (2009).

enal experiences. Yet, non-sensory phenomenal experiences also seem to

exist. This must be accounted for. I will explain why and how we represent some (but not all) non-sensory processes as phenomenal, focusing on three categories. First, mnemic and imaginative phenomenality. Second, cognitive and high-level perceptual phenomenality. Third, valenced phenomenality.

A. Mnemic and imaginative phenomenal experiences

Intuitively, when I bite into an apple, I have a *sensory* phenomenal experience of its taste. However, it seems that I can also have partially similar experiences when *remembering* or *imagining* biting into the apple—mnemic and imaginative taste experiences, whose phenomenal character seems close to the character of sensory experiences. Why and how do we introspect mnemic and imaginative processes as phenomenal? Why do we represent them as partially (but not entirely) phenomenally similar to some sensory experiences?

For Meta-Cartesianism, we monitor our non-inferential belief-formation processes and represent them as phenomenal. Unsurprisingly, we thus represent *some* non-sensory processes as phenomenal. This is the case with mnemic processes, which typically lead, non-inferentially, to beliefs—e.g., notably, though not only, beliefs about the past. Meta-Cartesianism also correctly predicts that varieties of memory that do not conduce to beliefs, like procedural memory, will not be introspected as phenomenal.

The case of imagination is more complex. First, one might doubt that imagination is a process of belief-formation. *Seeing* a tiger makes me believe there is a tiger, and *remembering* one makes me believe there was one, but I can *imagine* a tiger just for the sake of it. If imagination is not a belief-formation process, given Meta-Cartesianism, it should not be introspected as phenomenal. Yet, it is.

In response, it is uncontroversial that we *sometimes* form beliefs based on imagination, such as beliefs about what is possible or likely. Imagining a list of guests for a dinner party, I form beliefs about how the party would probably unfold. Imagining a red square, I realize it could exist; failing to imagine a round square, I conclude it could not. Going beyond this uncontroversial point, it is plausible that most, if not all, imaginative processes *dispose* us to form some beliefs, even if we do not *use* them to do so. I can imagine winning the lottery, not to learn about what would then likely happen, but just

because I enjoy the reverie. Still, doing so, I tend to form beliefs about what would happen then. I can imagine a tiger for the sake of it, but doing so, I still tend to form beliefs about what a tiger would look like.

So, imagination is a belief-formation process, but is it also *non-inferential*? The beliefs it makes us form seem to depend on what we already believe. Imagining guests at my dinner party, I conclude that they would fight about politics, but this follows from what I know about them.

Still, while beliefs shape imagination, what imagination disposes us to believe is not explicitly inferred from our beliefs through personal-level inferences. Imagination rather makes implicit beliefs and their implications manifest. This is why thought experiments are useful. Additionally, imaginative processes have their own internal dynamics, not entirely determined by preexisting beliefs. Imagining a purple tie worn with a green shirt, I conclude that these colors do not accord, something which none of my beliefs imply.

Therefore, it is plausible that imagination is a non-inferential process of belief-formation. Meta-Cartesianism thus correctly predicts that we will introspect it as phenomenal. It also correctly predicts that non-imaginative *suppositive processes* (where we suppose that P without imagining it, and see what follows) are not typically introspected as phenomenal, since these processes are inferential.

Finally, why are mnemic and imaginative experiences introspected as partially similar to some sensory experiences? Consider again the introspection of sensory experiences. Facing a red object, my sensory module outputs THERE IS RED; my ERM-powered Cartesian Device outputs THERE IS *RED* IN ME, *red* being estimated to resemble red distinctively and therefore to provide direct evidence for it.

What happens with mnemic and imaginative experiences? First, mnemic and imaginative processes produce certain outputs; maybe, THERE IS RED IN THE PAST for memory, and THERE IS RED IN THIS POSSIBLE SITUATION for imagination. Our ERM-powered Cartesian Device then outputs, respectively, THERE IS *RED IN THE PAST* IN ME, and THERE IS *RED IN THIS POSSIBLE SITUATION* IN ME.

These phenomenal experiences, which these outputs represent, are then estimated to *resemble* a sensory experience of red (as well as each other) but not perfectly so. Indeed, we use representational similarity as a proxy for experience similarity, and our representations of these experiences are all

[15] For the view that even non-sensory experiences are presentational, see Chudnoff (2012).

partially similar, since they all have *RED* as a component. Still, the similarity is not complete: their other components differ. So, all three experiences seem partially similar, as they all involve *red* and present red-involving situations to the subject.[15] This generalizes to other experiences.

It is also plausible that, as a matter of fact, sensory activity in the presence of the sensed object generates fine-grained and determinate outputs, while mnemic and imaginative activity generate coarser-grained, less determinate, more generic outputs. The tigers we *see* are usually more detailed than those we *remember* or *imagine*. Since these differently detailed outputs are redeployed in the introspection of these states, sensory phenomenal experiences will seem richer than mnemic and imaginative ones: we will grasp them as presenting richer, more fine-grained and determinate situations.

Earlier, I showed that Meta-Cartesianism explains our intuition that there is no distinctive phenomenology of introspection. We do not represent the way it feels to have an experience and the way it feels to have this experience introspectively appearing to us as distinct. Still, the view predicts a phenomenology of *imagining* or *remembering* the experiences of others. Imagining or remembering these experiences, I undergo a specific non-inferential belief-formation process, with outputs such as THERE IS *RED* IN HER IN THE PAST, or THERE IS *RED* IN HIM IN THIS POSSIBLE SITUATION. These processes are then *themselves* introspected as phenomenal: focusing on them as I have them, I produce introspective outputs like THERE IS *THERE BEING ⁎RED⁎ IN HER IN THE PAST* IN ME, or, THERE IS *THERE BEING ⁎RED⁎[16] IN HIM IN THIS POSSIBLE SITUATION* IN ME. The phenomenal character of these introspected experiences is then estimated to be partially similar to the character of other experiences represented with the concept *RED*, like simple sensory experiences of red. The same applies to non-inferential mindreading of phenomenal experiences, which should also be introspected as phenomenal.

As a result, much of our cognition about the phenomenal is introspected as phenomenal. Phenomenal introspection, as shown earlier, is introspected as itself phenomenal, but as I just showed, imagining experiences, remembering experiences, and representing them through non-inferential mindreading are also introspected as phenomenal.

[16] While I normally use *X* to refer to a phenomenal concept, I use ⁎X⁎ to refer to it when it is embedded within a larger one, to better describe thoughts about phenomenal experiences of imagining or remembering other phenomenal experiences.

[17] Papineau (2002); see also Loar (1990; 1997); Balog (2012a).

For Meta-Cartesianism, which is an illusionist view, our cognition about the phenomenal *never really* involves anything phenomenal. Still, we represent our phenomenal cognition as typically involving phenomenal experiences—the represented experience itself in the case of introspection, another experience partially resembling the represented experience in the other cases. This explains our intuition that we think of phenomenal experiences by way of phenomenal experiences themselves *or* "faint copies" of them.[17]

B. Cognitive and high-level perceptual phenomenal experiences

Other non-sensory processes, even more distant from the sensory, are introspected as phenomenal. Intuitively, facing a tomato, I have a sensory experience of a red, round object, but I might also have an experience of it *as a tomato*, or even *as a beefsteak tomato*. Some experiences seem to have high-level perceptual phenomenal features. Intuitively, we might also have *cognitive experiences*.[18] For instance, consider, beyond the mere sensory components, the experience of *understanding* a sentence,[19] a metacognitive feeling such as the tip-of-the-tongue feeling,[20] or an agentive experience of an action as being our own.[21]

Three things are notable about high-level perceptual and cognitive phenomenology. First, their existence is *debated*.[22] Some think that these experiences have a *sui generis* phenomenal character, which does not reduce to sensory, imaginative, or mnemic phenomenal character[23]; others deny it.[24] The debate is ancient: a version of it animated early 20th-century introspectionist psychology, when the Würzburg school, guided by Külpe and comprising researchers like Marbe, Ach, and Bühler, defended the existence of imageless thought against other prominent introspectionist psychologists like Wundt or Titchener.

[18] Some think that high-level perceptual phenomenology is nothing but a mix of sensory phenomenology and cognitive phenomenology (Montague 2017).
[19] Strawson (1994, 5–6).
[20] Goldman (1993).
[21] Synofzik et al. (2008).
[22] Bayne and Montague (2011).
[23] Strawson (1994); Siewert (1998); Pitt (2004); Kriegel (2015).
[24] Nelkin (1989); Tye (1995); Carruthers (2005a).
[25] Nichols and Stich (2003, 196).

Second, *some* cognitions and high-level perceptions are more plausible possessors of sui generis phenomenal character than others. Understanding a sentence in one's native language is a good candidate, but many doubt that judging that 19 is prime has a sui generis cognitive phenomenal character,[25] let alone an *individuative* one distinguishing it from judging that 17 is prime.

Third, these experiences do not seem qualitative or ineffable in the same way as sensory experiences.

Meta-Cartesianism is illusionist: it denies the *existence* of sui generis cognitive or high-level perceptual phenomenology. Still, it should explain their apparent existence, and the three facts I just mentioned, which count as three specific explananda.

Meta-Cartesianism explains that we sometimes introspect sui generis high-level perceptual and cognitive phenomenology. Suppose the corresponding processes sometimes constitute non-inferential belief-formation processes. This is something I find independently plausible. For instance, it is plausible that my metacognitive module sometimes outputs that *I know* something—say, the middle name of Joe Biden—even if I do not recall it at the moment. I do not explicitly infer this piece of metacognitive knowledge; I am just non-inferentially disposed to accept it. Similarly, it is plausible that my visual module sometimes outputs not only that there is a red round object in front of me, but also that it is *a tomato*, without it being explicitly inferred at the personal level. Given Meta-Cartesianism, these non-inferential belief-formation processes will be introspected as phenomenal.

The relevant outputs here are not formulated only with sensory concepts, like RED. They comprise high-level concepts (TOMATO) or abstract concepts (KNOW). Corresponding experiences are thus represented with phenomenal concepts like *KNOW* and *TOMATO*. These experiences will seem to have *sui generis* phenomenal character, as we represent them as having phenomenal properties that sensory experiences never have: e.g., *tomato* and *know*. They will also seem to have an *individuative* phenomenal character, as we represent them as having distinctive phenomenal properties—say, *tomato* versus *red pepper*. This explains the appearance of sui generis and individuative high-level perceptual and cognitive phenomenology.

High-level perceptual and cognitive phenomenal concepts are not formed by redeploying sensory concepts, but by redeploying high-level and abstract concepts. Those arguably do not work like sensory concepts: they do not place their referents in comparably rich quality spaces. Resulting

phenomenal concepts will inherit these features; high-level perceptual and cognitive phenomenal properties will not seem qualitative and ineffable exactly like sensory experiences. This explains our third explanandum.

Meta-Cartesianism also predicts that introspection will be less decisive about these phenomenal experiences than about sensory phenomenal experiences. We represent phenomenal experiences when an internal monitoring mechanism detects non-inferential belief-formation processes. This detection relies on various heuristics, using various cues. Plausibly, these cues are more ambiguous for high-level perceptual and cognitive processes than for sensory ones. To see why, consider four processes.

A: I see something round and red and infer, based on its sensory features, that it is a beefsteak tomato.
B: I see something and immediately recognize that it is a beefsteak tomato.
C: Using an artificial color sensor, I infer that there is something red.
D: I see, without inference, that there is something red.

A and B are about representing high-level features (inferentially for A, non-inferentially for B). C and D are about representing low-level sensory features (inferentially for C, non-inferentially for D). Interestingly, A and B seem functionally closer to each other than C and D. D differs from C by being clearly automatic and encapsulated. However, B is not as radically automatic and encapsulated when contrasted with A. There is a possible smooth transition between A and B: learning about tomatoes, I progressively deploy B where I used to deploy A. It is not clear that there is the same sort of smooth transition between C and D. Finally, in B, I typically undergo some of the mental processes involved in A. I *see* the object's basic sensory features, which *could* allow me to infer what I immediately see—that it is a beefsteak tomato. Nothing similar happens for C and D: seeing red, I do not usually also use an artificial sensor delivering information from which I could *infer* what I also see.

I will not argue for it, but plausibly, this generalizes: the cognitive profiles of *inferential* and *non-inferential* belief-formations about high-level or abstract features are more similar to each other than the profiles of inferential and non-inferential belief-formations about low-level sensory features. As a result, phenomenal introspection is less decisive about the first: it makes more ambiguous and less confident verdicts when introspecting high-level

[26] Baars (1988, 51).

perception and abstract cognition as phenomenal. We can also expect more interindividual differences in this case, downstream of how individual mechanisms' decision criteria are set. This explains why people debate the existence of high-level perceptual and cognitive phenomenology, not of sensory phenomenology. This was our first explanandum.

This also helps with our second explanandum. For Meta-Cartesianism, we represent a process as phenomenal when we detect it as a non-inferential belief-formation process. Modular processes, like sensory processes, typically qualify, but other processes have an ambiguous status. High-level perception and abstract cognition often start as non-modular, but become somewhat modular with practice. As cognitive psychologists have long noted:

> any highly practiced and automatic skills tend to become "modular"—unconscious, separate from other skills, and free from voluntary control.[26]

> even a post-perceptual decision process can, with time and repetition, become automatized and cognitively impenetrable, and therefore indistinguishable from the encapsulated visual system.[27]

This explains why the best candidate cases of high-level perceptual and cognitive phenomenology are deployments of expertise. Take expert object recognition: once I have seen a lot of tomatoes, I recognize beefsteak tomatoes at first sight, and it seems that I experience them as such. Before this, I had to infer a tomato's variety from its sensory features. Take expert sentence comprehension: once fluent in a language, I immediately understand a sentence, and it seems that I experience its meaning. Before this, I had to translate sentences based on fairly basic visual or phonetic features, using explicit knowledge of vocabulary and grammar.

So, Meta-Cartesianism explains our second explanandum. Expert routinized abstract processes are often introspected as sui generis experiences, but highly controlled processes relying on explicit inferences typically are not. This is why mathematical judgments (e.g., judging that 17 or 19 are prime) are not typically introspected as phenomenal, even if things might be different with expert mathematicians, who may introspect mathematical feelings and say, e.g., that a number "feels prime."[28]

[27] Pylyshyn (1999, 360).
[28] Kendall (1968, 159). The mathematician cited is Alexander Aitken.

C. Valenced phenomenal experiences

We introspect phenomenal experiences with affective components. These are valenced experiences. Some are *attractive* and positively valenced, like pleasure, joy, and euphoria: those feel good. Some are *aversive* and negatively valenced, like pain, anger, and fear: those feel bad. By contrast, many sensory experiences, such as seeing a gray circle, apparently lack valence. They seem neither aversive nor attractive, but neutral. *Valence* seems a part of the phenomenal character of valenced experiences: intuitively, if a typical experience of pain were deprived of its aversive character and did not feel bad, it would be phenomenally different.

Meta-Cartesianism is illusionist. It denies phenomenal valence. Still, it must explain its appearance. Why do we introspect valenced phenomenal experiences? Why does valenced phenomenology seem distinctive?

To explain this, I first need assumptions about *affective cognition*: the actual mental processes responsible for pain, pleasure, joy, sorrow, etc. I assume a form of *evaluative representationalism*.

Remember: I assumed that *sensory perception* relies on sensory modules producing outputs formulated with sensory concepts; sensory concepts track worldly properties, which they misrepresent as primitive sensory qualities. Evaluative representationalism adapts this account to affective cognition, with a twist to explain valence. Affective cognition is *modular*—although it might not correspond to a distinct module, but rather be intertwined with sensory modules. It tracks certain features of events, like bodily events: their being fitness-improving or fitness-decreasing. It represents these features as instantiations of primitive value: as primitively *good* or *bad*.

For instance, in the case of pain, modular affective processes, intertwined with bodily sensory perception, track fitness-decreasing bodily events and represent them as bad. In pleasure, they track fitness-increasing bodily events and represent them as good. Doing so, they deploy value concepts: GOOD and BAD. Like sensory concepts, these value concepts characterize their referents

[29] For a review of theoretical options focusing on the case of pain, see Aydede (2019). My view corresponds to what he calls the "mixed theory." See also Coninx (2024).

[30] Peter Carruthers (2018, 2023) might be the most prominent current defender of the idea that valenced states are representations of value, over the imperativist alternative, on which they have imperative content. Carruthers views these representations as non-conceptual.

[31] The idea that pain has a sensory and an affective component is standard (Aydede 2019; Coninx 2024).

as primitive. They gain their peculiar axiological content in virtue of their implicit functional links to a range of conative and cognitive reactions: e.g., BAD triggers conative reactions (desiring that bad things stop, etc.) and cognitive reactions (judging that bad things *should stop*, etc.).

For evaluative representationalism, affective cognition is a matter of representing the instantiation of values. Evaluative representationalism is not the only game in town about affective cognition,[29] but it is influential and reasonably plausible.[30]

Here are a few more examples analyzed through this variety of evaluative representationalism. Burning my hand on a hot stove, affective cognition, intertwined with sensory perception, produces an affective-sensory output: THIS IS BAD; THIS is a bodily concept referring to something happening in my hand, and BAD is a concept of primitive negative value.[31] When I drink a glass of a good Saint-Estèphe, affective cognition, intertwined with sensory perceptions, outputs an affective-sensory verdict: THIS IS GOOD. THIS, here, is a bodily concept referring to something happening in my mouth, and GOOD is a concept of primitive positive value. Waking up on a sunny summer day, I am joyful: I output an affective-abstract verdict, THIS IS GOOD; THIS is an abstract concept referring to my current situation, and GOOD is the concept of primitive positive value.

These affective outputs are non-inferential belief-formation processes. They non-inferentially lead to beliefs about what is good and bad. Given Meta-Cartesianism, they are introspected as phenomenal. In the three described cases, I introspect phenomenal pain, pleasure, and joy, redeploying GOOD and BAD to form and apply *GOOD* and *BAD*, which are phenomenal concepts representing phenomenal goodness and badness.

This explains why we introspect valenced phenomenal experiences and why valenced phenomenology stands out: we represent it as involving primitive phenomenal values absent in neutral experiences. We represent valenced phenomenal experiences as uniquely providing direct evidence for things being good or bad.

A side note: given the tight conceptual links between the goodness and badness of things and the fact that they should or should not happen, we also view valenced phenomenal experiences as uniquely giving direct evidence that things should or should not happen and, ultimately, non-inferential reasons to do certain things.

6.6. Explaining our intuitions of significance

A. Epistemic significance

Finally, Meta-Cartesianism explains our significance intuitions. Start with our intuition of epistemic significance. For Meta-Cartesianism, we grasp the phenomenal as essentially giving direct evidence: as a medium through which non-phenomenal reality epistemically appears. We also represent it as introspectively self-revealing and infallibly introspectively appearing: as standing in a unique epistemic relation to us. All this stems from the content of our phenomenal concepts.

Therefore, Meta-Cartesianism explains a shared, compelling, and conceptual intuition that phenomenal consciousness plays an epistemic role. It explains an intuition of epistemic significance.

Besides, from the perspective of our naïve evidential cognition, phenomenal experiences are the only source of direct evidence. Therefore, they are the only possible stopping points in the chains of indirect evidence we represent. This naturally leads to the idea that the phenomenal is not only epistemically significant, but also epistemically unique and distinctive, if not epistemically indispensable. For more on this and on illusionism's epistemological implications, see Chapter 10.

B. Ethical significance

Phenomenal experiences also seem ethically significant. Valenced phenomenal experiences, notably, seem to play a distinctive role in the well-being of beings who have them. The capacity to have those experiences—*sentience*, in the narrow sense of the term—seems to play a distinctive role for moral status. Intuitively, sentient creatures, because they are sentient, deserve a distinctive sort of moral consideration.

In Chapter 1, I noted that intuitions regarding well-being and moral status are probably connected. Plausibly, welfare subjectivity—the property of having a certain well-being—is necessary and sufficient for moral status, and the *degree* of well-being one can reach impacts one's moral status. Explaining our intuition that the phenomenal is prudentially significant helps explain the intuition that it is morally significant.

Meta-Cartesianism explains our intuition of prudential significance. First, consider affective cognition without introspection: it modularly represents things as good or bad, using concepts of primitive value—GOOD and BAD. This is what happens in pain, pleasure, joy, sadness, etc. Plausibly, there is a tight conceptual relation between prudential value and the primitive value represented in affective cognition: pain, pleasure, and comparable states make us represent things as *prudentially* valuable. For instance, in gustative pleasure, I represent something happening in my mouth as good, but also as *good for me*, as contributing positively to my well-being.

We also introspect affective cognition as involving valenced phenomenal experiences. We represent these experiences as instantiating phenomenal goodness and badness, grasped with GOOD and BAD: phenomenal versions of the primitive non-phenomenal values that affective cognition represents, which distinctively resemble them, and give direct evidence for them. This plausibly explains why we think of valenced phenomenal experiences as contributing to well-being. The tight conceptual relation between prudential value and the primitive value affective cognition represents translates into a tight connection between prudential value and the resembling primitive phenomenal value phenomenal cognition represents.

This explains a shared, compelling, and conceptual intuition that valenced phenomenal experiences contribute to well-being. Still, it does explain why they seem *distinctively, if not exclusively*, prudentially valuable.

To do so, come back to our treatment of irreducibility intuitions. Many things—those represented as primitive—seem irreducible. We usually discard these intuitions as irrelevant and bearing merely on apparent and non-existent primitive properties, but we cannot do this for phenomenal properties, which we grasp as introspectively obvious. If we accept the relocation story told earlier, we first believed in a world of primitive sensory qualities, but Galileo and Descartes taught us that those are not real. We relocated primitive qualities in ourselves, in the phenomenal world, where everything is as it appears. We denied primitive sensory qualities, but conceded primitive *phenomenal* properties.

A *normative* relocation story is also plausible. Affective cognition represents external things and bodily events as good or bad for us. We initially took it at

[32] O'Keefe (2016, 29).
[33] In the *Anti-Seneca or the Sovereign Good*, first published in 1748 (La Mettrie, 1996).
[34] *"The question is not, Can they reason?, nor Can they talk? but, Can they suffer?"* (Bentham [1789] 1961).

face value and believed in a non-phenomenal world containing good and bad things. We believed that what happens in our shoulders during massages is intrinsically good for us, while what happens in our hands when we burn them is intrinsically bad for us. We then came to admit that primitive value does not really inhere in external things and bodily events. Maybe we failed to find a place for them in a physical world; maybe we found it too hard to rationalize contradictory value judgments about similar external items. We concluded that nothing non-phenomenal is, in itself, good or bad for us. The value was relocated in phenomenal experiences, where things are as they seem.

The heroes of this normative relocation were not Galileo and Descartes. In the Ancient World, they were Epicureans and the Cyrenaics, who thought pleasure was the only final good; or Democritus, who "denied that pleasure was the good, [but] did identify the good with a state of mind: [. . .] cheerfulness."[32] In the modern period, they were, perhaps, La Mettrie, who considered pleasure the only ultimate good,[33] or Bentham,[34] who thought pleasure was the only good, pain the only evil, and viewed narrow sentience as the criterion of moral status.[35]

Post-relocation, we thought that anything good or bad for us is a matter of phenomenal experiences. Having *emptied* the non-phenomenal world of its value, phenomenal experiences were left as the only locus of value, with the non-phenomenal having at most instrumental value, when it brings about the right experiences.

This normative relocation story is simplistic, but there is a core of truth to it. It explains our impression that phenomenal experiences are not only prudentially valuable, but also distinctively and even exclusively so. Assuming a connection between welfare subjectivity and moral status, it also explains why sentience seems to be distinctively and even exclusively important for moral status. Note that, given this explanation, these intuitions are not widely

[35] The normative relocation and the standard relocation have different timelines. Why? On the one hand, there might be more disagreement about non-phenomenal value than about non-phenomenal sensory quality: e.g., we disagree more about external good and bad things than about external colors. This makes the normative relocation more attractive. This might explain why it was popular earlier on, in the Ancient World. On the other hand, the normative relocation might go against Christian beliefs, which would explain why, in the West, the secondary quality relocation became accepted after the Scientific Revolution of the 17th century, while the normative relocation only became popular with the secularization starting in the 18th century.

[36] Chalmers (2022).

[37] "There are no perceptions which are matters of complete indifference" (Leibniz [1765] 1996, p. 162).

shared, compelling, nor conceptual, as they depend on the antecedent acceptance of the normative relocation.

Finally, a worry subsists. All this explains why *valenced* phenomenal experiences and *narrow* sentience seem ethically significant, but arguably, we view *neutral* phenomenal experiences as ethically significant too. *Vulcans*—imaginary creatures capable only of neutral phenomenology—seem to some to have welfare and moral status. Can we explain this intuition?[36]

I see two strategies to do so within the framework I proposed. First, we could maintain that we never really represent neutral phenomenal experiences: we grasp all experiences as valenced, even if just a bit. This would explain why all phenomenal experiences seem ethically significant. The idea that all phenomenal experiences are represented as valenced is not ad hoc, since many philosophers independently thought that all experiences are indeed valenced. Leibniz thought so and argued for the view against Locke,[37] who himself believed that almost all actual experiences are valenced.[38] Arguably, it was the opinion of Aristotle.[39] Empirical researchers also recently argued that apparently "neutral" visual perceptions are loaded with micro-valences.[40]

Second, we could suppose that our significance intuition about phenomenality in general, including neutral experiences, has another source. Phenomenality seems irreducible: this might render phenomenal consciousness "sublime,"[41] and aesthetically valuable. The phenomenal appears to consist of a myriad of primitive, irreducible qualities and thus seems extraordinarily diverse compared to a physical world exhausted by instantiations of mathematical and causal structures. This plausibly makes it seem valuable.

Both explanations are plausible to some extent, and both correctly predict that significance intuitions about neutral (or apparently neutral) phenomenal experiences will be weaker than those about valenced experiences.

[38] "*Delight*, or *Uneasiness*, one or other of them join themselves to almost all our *Ideas*, both of Sensation and Reflection" (Locke [1689] 1975, 128).

[39] This is an interpretation of Aristotle's claims, in the *Metaphysics*, about "our liking for the perceptual capacities" which, "even apart from their utility, [are] liked because of themselves" (Aristotle 2016, 2). It follows from viewing pleasure as "unimpeded activity of a natural state," including activity of the "sense faculties" (Kraut 2022).

[40] Lebrecht et al. (2012); see also Barrett and Bar (2009); Cleeremans and Tallon-Baudry (2022).

[41] Niikawa and Kriegel (2025).

6.7. Objections to Meta-Cartesianism

To close this chapter, I discuss specific objections to Meta-Cartesianism. Objections to illusionism in general are treated in the next chapter.

First, one might worry that Meta-Cartesianism is too speculative. Against this, note that the view gains plausibility precisely because it solves the illusion problem. Moreover, while speculative, the view makes confirmed or testable predictions beyond the illusion problem. It correctly predicts that resemblance views of perception will be deeply intuitive. It predicts that manipulating the inferential nature of a process would alter its introspection as phenomenal—a testable prediction.

Besides, remember that the reason why we want to solve the illusion problem is not just that the appearance of phenomenal consciousness is among our explananda. It is also that, given the dialectical situation, illusionists must show that, in principle at least, the appearance of phenomenal consciousness is explainable without appealing to phenomenal experiences. Even if Meta-Cartesianism is highly speculative and likely to be incorrect at some level, it still gives a 'how-possibly' explanation of the illusion of phenomenal consciousness, on which this illusion is uniquely compelling, making realism false but not highly blamable. This explanation brings corresponding dialectical benefits.

Second, specific objections were formulated against some ancestors of Meta-Cartesianism and might apply to it. Regarding an earlier view of mine, close to Meta-Cartesianism in some respects, David Chalmers objected that, if it explains why we resist thinking of phenomenal consciousness as illusory, it does not explain why we resist views simply stating that phenomenal consciousness *does not exist*.[42]

Here is my response. Meta-Cartesianism predicts that we struggle to represent *illusions* of phenomenal experiences, and we do. It predicts that we do not struggle to represent the mere *inexistence* of phenomenal experiences, and we do not: we easily represent worlds without phenomenal experiences, such as zombie worlds or the early universe. Of course, Chalmers aims at something else. While zombie worlds are easy to represent, we still struggle to envision the *actual* absence of phenomenal experiences—their absence in *our* world—even if we do not explicitly add that we have an illusion of those. Meta-Cartesianism explains this, even if it does so by ultimately appealing to our difficulty to

[42] Chalmers (2018, n. 19).

represent illusions of phenomenal experiences. When we envision the *actual* absence of phenomenal experiences, we envision a hypothesis contradicting some of our strong and robust beliefs, as well as continuous and robust introspective outputs. For Meta-Cartesianism, given the functioning of our innate evidential cognition, to make sense of false strong and robust beliefs, as well as the deceptive robust output of non-inferential belief-formation capacities, we represent that we are being misled by deceptive direct evidence. When we consider that phenomenal experiences might be non-existent, given our beliefs and introspective outputs, we almost irresistibly tend to think that these phenomenal experiences must at least *seem* to exist, in the sense of our intuitive grasp of direct evidence: we must at least have an *illusion* of phenomenal experiences. This raises the difficulty explained above. To avoid this, we must simultaneously repudiate our previous strong and robust beliefs, discard robust introspective outputs, and suspend our intuitive manner of making sense of these incorrect beliefs and outputs with our concept of deceptive direct evidence. This is simply very hard to do.

Against a more recent ancestor of Meta-Cartesianism positing an ERM,[43] Chalmers claims that he is "doubtful about resemblance theories of representations."[44] This is no objection: Meta-Cartesianism entails that resemblance views are intuitive, but false. Chalmers adds that he is "doubtful that even our cognitive systems are committed to [resemblance theories]." This is an objection. Still, Meta-Cartesianism does not entail that evidential cognition explicitly encodes a resemblance theory of perception or representations, only that we use a measure of resemblance to represent what phenomenal experiences give direct evidence for. This correctly predicts that resemblance views of perception should be very intuitive, not that the views should already be explicitly formulated or believed by everyone. Meta-Cartesianism only implies that our cognitive systems are 'committed' to resemblance theories of perception in a highly limited and qualified sense.

Chalmers also points out that this ancestor of Meta-Cartesianism fails to explain why cognitive states like beliefs do not generate the same problematic intuitions as conscious perceptions. Josh Weisberg later pressed a similar worry against my view: Why is it that a sensation of red is introspected as phenomenal, with all the related intuitions, while the thought that 2 + 2 = 4 is not?[45]

[43] Kammerer (2019c).
[44] Chalmers (2018); Chalmers (2020b, 206).
[45] Weisberg (2023, 80–81).

The previous section on high-level and cognitive phenomenology helps address this. Many abstract thoughts are not introspected as phenomenal, because they are not monitored as appropriately non-inferential: e.g., the thought that 2 + 2 = 4 is typically embedded in an inferential network of mathematical beliefs, in a way in which a sensation of red is not. Meta-Cartesianism explains why *some* abstract cognitive processes are introspected as phenomenal, while correctly predicting that introspection is typically more ambiguous and less decisive about them than about sensations. It explains why even these processes do not seem qualitative and ineffable, which implies that, even when we introspect these states as phenomenal, they still seem pretty different from the paradigmatic phenomenal experiences—phenomenal sensations, and related phenomenal states.

* * *

Meta-Cartesianism solves the illusion problem. It explains the illusion of sensory phenomenal experiences. It explains our mystery intuitions. It accounts for the illusion of non-sensory phenomenal experiences. It explains our intuitions of significance. Specific objections to the view can be answered. In the next chapter, I address objections to illusionism in general.

7
Addressing Objections

Some people [. . .] contend that nothing prevents us from supposing that the heavens remain motionless while the earth revolves on its axis, with approximately one revolution a day [. . .]. However, they do not realize that [. . .] such a notion is ridiculous.

—Claudius Ptolemy

In this chapter, I answer eight objections to illusionism. First, the objection that illusionism is incoherent, because phenomenal consciousness cannot seem to exist without existing. Second, the objection that illusionism cannot explain how we represent phenomenal properties. Third, the objection that illusionism is only attractive if we assume a heavily loaded definition of "phenomenal consciousness." Fourth, the objection that it is only attractive if mystery intuitions are widely shared, but they are not. Fifth, the objection that it is only attractive if mystery intuitions are conceptual, but they are not. Sixth, the objection that it is only attractive if mystery intuitions correspond to ineliminable commitments of phenomenal concepts, but they do not. Seventh, the objection that illusionism, even if coherent and argumentatively supported, is obviously false. Eighth, the verbal objection that strong illusionist formulations are inferior to substantively equivalent weak illusionist (and realist) formulations. I conclude by examining miscellaneous worries left unaddressed. Objections appealing to normative data, regarding epistemic or ethical facts, are discussed later, in Chapters 9 and 10.

7.1. The no-appearance/reality-gap objection

Many object to illusionism that it is incoherent, as the *appearance* of phenomenal consciousness implies its *existence*. This is supported by the intuitive idea that, regarding the phenomenal, there is no gap between appearance

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0008

and reality. This objection has been raised, for instance, by John Searle and Galen Strawson:

> You can't disprove the existence of conscious experiences by proving that they are only an appearance disguising the underlying reality, because *where consciousness is concerned the existence of the appearance is the reality.*[1]

> What is it to suppose that one might be completely wrong [regarding the existence of one's phenomenal experiences]? It is to suppose that although it *seems* to one that there is experience—for this cannot be denied—there really isn't any experience. But this is an immediate reductio ad absurdum. For this seeming is already experience.[2]

The objection is easy to answer if we distinguish various senses of "appearance," as I did earlier.[3] Phenomenal consciousness cannot *phenomenally* appear to us without existing, but this is a cheap conceptual truth, which does not bother illusionists, who think that nothing phenomenally appears. Phenomenal consciousness *psychologically* appears to us to exist without existing: we are disposed to think it exists, even if it does not. However, there is no reason to think that gaps between *psychological* appearance and reality are impossible regarding the phenomenal. It also *epistemically* appears to exist without existing: we have some reasons to believe that it exists, even if it does not. However, again, there is no incoherence there.

As noted previously when discussing the illusion meta-problem, these last two claims are coherent, but they still generate intuitive tensions. Evidential cognition, based on our Cartesian Device, represents—incorrectly—that direct *epistemic* appearances are phenomenal appearances. As a result, we intuit that, if a phenomenal experience directly epistemically appears to exist, it phenomenally appears to exist, which implies that something phenomenal exists. Evidential cognition also represents—incorrectly—that direct epistemic appearances of one's phenomenal experiences cannot be deceptive. Consequently, we intuit that, if a phenomenal experience directly epistemically appears to exist, it exists. So, the claim that phenomenal consciousness directly epistemically appears to exist without existing generates deep intuitive tensions. The same applies to the claim that it psychologically seems to

[1] Searle (1997, 112).
[2] Strawson (1994, 51–52).
[3] See also Frankish (2016, 32–34).

exist without existing, since, as I argued previously, we intuit that this psychological appearance must come with a direct epistemic appearance.

Illusionists deny that phenomenal consciousness exists. They also deny that direct epistemic appearances are grounded in the phenomenal *and* that direct epistemic appearances of one's own phenomenal experiences are infallible. All these claims are intuitive: phenomenal and evidential cognition are intertwined. Both are mischaracterizing. The no-appearance/reality-gap objection is misguided, but it will always remain intuitively attractive.

7.2. The phenomenal representation objection

Another objection, advanced by Joseph Levine,[4] concerns illusionism's capacity to explain phenomenal representations. For illusionists, there are no phenomenal properties. Still, we *represent* them, and not as mere *je-ne-sais-quoi*: our phenomenal concepts give us a *substantive grasp* of phenomenal experiences. As Levine puts it:

> When I think of what it is to be [phenomenally red], [phenomenal redness] is somehow included in the thought: it's present to me. This is what I mean by saying it has a "substantive" mode of presentation. In fact, it seems the right way to look at it is that [phenomenal redness] itself is serving as its own mode of presentation.[5]

Brian Loar made a similar claim:

> [A] phenomenal concept has as its mode of presentation the very phenomenal quality it picks out.[6]

Suppose that phenomenal properties do not exist. Where do we get our representations of them? The phenomenal representation objection is that

[4] Levine (2001, 143–49). See Frankish (2016, 35–37) for a discussion. The core idea of the objection is already present in Sydney Shoemaker's remarks about projectivism (Shoemaker 1990, 128). For other discussions of similar points, see Rey (2007); Robinson (2022); Shabasson (2022); Duncan (2023).

[5] Levine (2001, 8).

[6] Loar (1997, 604). The claim is only plausible if interpreted liberally, so that the included phenomenal property can be a mere faint version of the property thought about—e.g., we can think about excruciating pain without experiencing it fully.

we cannot satisfyingly answer the question. Representing phenomenal properties requires instantiating them. Illusionism is false.

To show that this question has no satisfying answer, the objector can first note that many theories of content require, for a mental representation to represent Fs, that this representation (or one of the same type) is or was in certain *causal* or *informational* relations with Fs. Plausibly, this is true for informational semantics,[7] or teleosemantics,[8] which require that representations of red have (or had) appropriate causal or informational relations with red surfaces. If phenomenal properties do not exist, they never had such relations to any of our mental states. How could we represent them?[9]

In response, illusionists can appeal to views of content that do not require actual links between the representation and the represented,[10] such as functional-role (or conceptual-role) semantics.[11] Given functional-role semantics, a representation's content is grounded in the functional role of the representation, including its causal links to other representations and behavior, without the need for actual relations with the represented. Functional-role semantics could provide (part of) the right account of phenomenal concepts' content. Of course, illusionists do not need to accept functional-role semantics across the board.

Levine concedes that we represent various non-existent entities, like Santa Claus, or pink elephants, in virtue of the functional role of the relevant representations. We can think of Santa Claus because our representation of him has the right functional links to representations of Christmas, old men, beards, red, reindeer, etc. In my view—Meta-Cartesianism—phenomenal concepts similarly refer to non-existent phenomenal properties in virtue of their functional links to concepts of evidence, agents, primitivity, sensory concepts, etc.

However, Levine stresses a disanalogy between phenomenal properties and Santa Claus, or pink elephants. These last entities do not exist, but the properties by way of which we think of them are all instantiated, even if never together in the right way. Men, beards, elephants, red, and pink

[7] Stampe (1977); Dretske (1981); Fodor (1987).
[8] Millikan (1984); Papineau (1984); Dretske (1988; 1995).
[9] Another way of making the same point is this: illusionism implies that we reliably misrepresent certain non-phenomenal states as phenomenal. However, the theories of content I just mentioned, which are all "tracking" theories, might not allow for reliable misrepresentations (Mendelovici 2013).
[10] Frankish (2016, 36).
[11] Harman (1973; 1987); Loar (1981); Block (1986).

exist. In contrast, phenomenal properties do not seem to consist of complex properties whose basic components are individually instantiated.

Does this disanalogy create a problem for our explanation of phenomenal representations? No. Suppose, for the sake of the argument, that representations of non-existent entities *must* be representations of complex properties, whose components are individually instantiated (though never together in the right way). Given Meta-Cartesianism, phenomenal concepts satisfy this condition. My concept of phenomenal redness characterizes its referent as a primitive property instantiated by a subject, which distinctively resembles red—represented as a primitive quality of surfaces—and gives the subject direct evidence for the presence of red, this direct evidence being itself characterized as introspectively self-revealing and infallibly introspectively presented. Plausibly, most basic features mobilized here—resemblance, primitivity, evidence, independence, beliefs—are instantiated in reality. *Red* as a primitive quality of surfaces is not, but basic features required to think about primitive red—primitivity, locations in multi-dimensional quality spaces, etc.—arguably are.

A harmless disanalogy remains between the consciousness case and the cases of Santa Claus or pink elephants. The links between phenomenal concepts and these other representations, in virtue of which phenomenal concepts have their content, are implicit. They depend on sub-personal mechanisms and are neither obvious nor immediately accessible to the subject. By contrast, links between concepts of Santa Claus and representations of beards, men, white, etc., are explicit. They depend on personal-level composition and are straightforwardly accessible to the subject. This is why, intuitively, being Santa Claus is nothing but instantiating the right combination of instantiated properties, while the same is not intuitive for phenomenal properties. However, this is unsurprising if an encapsulated module governs phenomenal concepts.

This explains how phenomenal concepts represent uninstantiated phenomenal properties. The objector, of course, may insist that this does not explain why these concepts are *substantive* in Levine's sense: why they make phenomenal properties present to us and included in our thoughts. However, illusionists should deny this extra explanandum. We *do not* have substantive representations of phenomenal properties in this strong sense. Still, Meta-Cartesianism explains why it seems to us that we do. As noted earlier, phenomenal cognition intuitively involves something phenomenal. Phenomenal introspection seems to involve the introspected phenomenal experience

itself. Imagining, remembering, or non-inferentially representing others' experiences seems to involve faint versions of the represented experience.

So, phenomenal cognition typically seems to involve phenomenal experiences, even if, according to illusionists, it does not. This is why, intuitively, our thoughts about the phenomenal seem substantive in Levine's sense—why they seem to include phenomenal properties. This also explains why an objection rebuked earlier, in Chapter 2, is intuitive: the objection that illusionism is a priori false, because merely *thinking* about phenomenal consciousness puts us in a position to know it exists. Phenomenal cognition typically seems to involve phenomenal experiences. However, we characterize phenomenal experiences as self-revealing. Consequently, it seems to us that *thinking* about phenomenal consciousness typically provides us with infallible direct evidence that we have some phenomenal experience, and thus gives us sufficient evidence to rule out illusionism. Therefore, illusionism seems a priori false. However, in reality, phenomenal cognition never involves phenomenal properties. Illusionism is a posteriori true.

7.3. The definition objection

Turn now to a third objection.[12] To argue for the eliminative claim in Chapter 3, I used a tandem of arguments, including the *Argument from Mysteries*. For the argument to succeed, one must admit that, as phenomenal consciousness seems mysterious, accepting its existence is problematic for attractive views—physicalism, or the view that our best sciences of the mind can in principle explain mental phenomena, or the view that reality can be studied objectively. Someone who denies that accepting the existence of phenomenality is problematic in this sense can reject the argument.

Suppose, for the sake of the discussion, that eliminativism is attractive if and only if accepting the existence of phenomenal consciousness is *problematic* in this sense.[13] An objection arises. It seems that accepting the existence of phenomenal consciousness is problematic only if we first define

[12] For a first version of my treatment of this objection, see Kammerer (2025).

[13] Traditional illusionist arguments presuppose that phenomenality is problematic (Dennett 1988; Frankish 2016). I am skeptical of arguments (Irvine and Sprevak 2020, 354–59) that avoid this presupposition. Chalmers's debunking argument for illusionism (Chalmers 2018, 44–49) does not presuppose that phenomenality is problematic, but it only supports a disjunction of weak and strong illusionism. To get to strong illusionism, Chalmers (2018, 49–52) presupposes that accepting phenomenality leads to the hard problem—i.e., that accepting its existence is problematic.

"phenomenal consciousness" in a loaded manner, stipulating its problematic nature.

Some illusionists defined it in this way. Dennett stipulated that "qualia" refers to properties that are ineffable, intrinsic, private, and immediately apprehensible in introspection.[14] One could add introspectively obvious, irreducible, etc. If we define "phenomenal consciousness" in this way, in terms of problematic features, accepting the existence of phenomenal consciousness is problematic, and eliminativism is attractive. However, the objection goes, this is not how realists define "phenomenal consciousness." Some realists *believe* that phenomenal properties have problematic features, but not all do, and even those who believe it do not make it a matter of definition. Realists use *neutral* definitions. They focus on examples of phenomenal experiences and invite us to focus on what they have in common, with the help of expressions such as "what it is like." They do not mention problematic features like introspective obviousness or non-physicality at the definitional stage.

This is the definition objection: if we define "phenomenal consciousness" neutrally, as realists do, accepting the existence of phenomenal consciousness is not problematic, and eliminativism ceases to be attractive.

Another way of stating the objection stresses that eliminativism is attractive if and only if phenomenal consciousness *conceptually* seems mysterious. If (and only if) problematic characterizations follow from phenomenal concepts, it is attractive to deny that these concepts are satisfied and that phenomenal experiences exist. This does not happen if our mystery intuitions do not stem from the way our concepts characterize their referents: if so, what is attractive is to maintain that these concepts are satisfied even if our mystery intuitions are incorrect. The definition objection contends that the concept of phenomenality used by realists is neutral and does not problematically characterize its referent, contrary to the loaded concept used by illusionists.

The objection is influential. Dennett discussed it and associated it with remarks made by Shoemaker.[15] Joseph Levine raised a similar point when distinguishing bold and modest qualiophilia.[16] Many realists insisted that they define "phenomenal consciousness" without appealing to problematic

[14] Dennett (1988, 47). See also Frankish (2016).
[15] Dennett (1988, 47–48).
[16] Levine (1994).

features.[17] An early version of the objection even appeared in the 1930s, directed at behaviorism.[18]

Illusionists standardly answered the objection by claiming that the loaded definition of phenomenality is the one everyone uses. For Dennett, when realists try to define phenomenal consciousness *without* appealing to problematic features, they fail to articulate their definition. This suggests that their mention of a neutral definition of "phenomenal" (or "qualia") is just "a gesture in the direction leading back to ineffable, private, directly apprehensible ways things seem to one."[19]

There is a core of truth in this illusionist response, but it remains unsatisfying, for three reasons. First, it seems uncharitable to suppose that realists do not know what they mean or are insincere. Second, the realists' failure to articulate an explicit neutral definition of consciousness does not reliably indicate that they use a loaded one. We often employ concepts whose definitions we cannot explicitly articulate, such as the concept of knowledge, as suggested by the failure of post-Gettier analyses. We could use a contentful and neutral concept of phenomenality even if we cannot explicitly articulate its definition. Third, this illusionist response is inconvenient for my version of illusionism. If the commonly used concept of phenomenal consciousness is explicitly loaded, it is hard to maintain, as I did, that the appearance of phenomenality is a widely shared and compelling illusion. Indeed, it is implausible that most non-philosophers tend to believe in phenomenal experiences if those are explicitly defined as entities that are ineffable, irreducible, etc. Supposing that the commonly used concept of phenomenality is this explicitly loaded one makes "meta-illusionism" attractive[20]: a version of theoretical eliminativism, in which there are no phenomenal properties but also no *illusions* of them. On this view, phenomenal properties do not exist, but they also do not widely appear to exist, and only a few highly sophisticated philosophers think they exist.

To answer the definition objection, go back to the definition of phenomenal consciousness I gave in Chapter 1. Like realists, I defined "phenomenal consciousness" by examples, directing introspective attention with the help of expressions such as "what it's like," "feelings," etc. The mysteries of

[17] Carruthers (2000); Kind (2001); Tye (2002). The citations are taken from Irvine and Sprevak (2020, 354). See also the examples reviewed in Frankish (2012, 669–73), and the argument made in Niikawa (2021).
[18] Williams (1934).
[19] Dennett (1988, 48). Frankish (2012) concurred.
[20] Mandik (2016).

consciousness were not part of the definition. These mysteries were definitional props among others. I did *not* stipulate that phenomenal consciousness is problematic. Instead, I claimed that the concepts we deploy when we appropriately direct our introspective attention to the examples given are phenomenal concepts, which happen to characterize their referents as problematic. Mystery intuitions could then serve as definitional props to help us latch onto our phenomenal concepts, even if the definition does not stipulate that phenomenal consciousness is mysterious.

There is no contradiction between the claim that our definition of phenomenal consciousness is explicitly neutral *qua* definition *and* the claim that phenomenal concepts characterize their referents as problematic. The apparent tension disappears when we understand that this problematic characterization is *implicit*. It is borne by phenomenal concepts, but it is not the case that it can be articulated by any competent user of these concepts, nor that the *negation* of problematic features corresponding to this characterization yields formal contradictions. Our definition of "phenomenal consciousness" does not explicitly make it a problematic entity, but it makes us latch onto a concept that implicitly characterizes its referent as problematic.

This line of reasoning is unsurprising to those familiar with debates on free will. Some think that the view that there is no free will is attractive if and only if free will is *incompatible* with determinism. Philosophers then debate whether our shared concept of free will characterizes its referent as incompatible with determinism. Since the compatibilist view, on which free will is compatible with determinism, does not seem contradictory, the characterization can only be implicit. This putative characterization is supposedly revealed by patterns of intuitions about cases, which become the topic of debate: Are we "natural compatibilists"[21] or "natural incompatibilists"?[22]

Another way of presenting my response is this. The definition objection stresses that we can deny the existence of mysterious mental states (i.e., obvious, irreducible, and undetectable states) without having to deny the reality of *this*, where "this" expresses THIS, a phenomenal concept. My response is that, if I am right about phenomenal concepts, THIS grasps its referent in a substantive way, which makes it mysterious, out of implicit conceptual necessity. As a result, one cannot coherently deny that correspondingly mysterious mental states exist while maintaining that *this* exists. We

[21] Nahmias et al. (2005); Murray and Nahmias (2014); Cova (2023).
[22] Nichols and Knobe (2007); Nadelhoffer (2020).

can *say* such things, but in this case, either we say something implicitly incoherent, or our use of "this" does not express a phenomenal concept, but some other non-phenomenal concept—see my discussion of the *redefinition* objection below.

Dennett's response to the definition objection has a core of truth: the concept of the phenomenal everyone uses makes phenomenality problematic, and this renders eliminativism attractive. Still, we must complete this response. The concept everyone uses is explicitly neutral, but it is *implicitly* problematic. Its problematic characterization does not stem from an explicit definition but from the sub-personal, inflexible, and innate functioning of our metacognitive module.

7.4. The rarity of mystery intuitions objection

Suppose phenomenal concepts characterize phenomenal experiences as mysterious. Suppose also that the appearance of phenomenal experiences is widely shared. This implies that mystery intuitions are widely shared. An objection is that they are not.

The objection can appeal to the history of thought. Historically, mystery intuition reports are rare and limited to certain periods and areas. This is notably the case for irreducibility intuitions. This issue was addressed in Chapter 1, where the scarcity of irreducibility intuition reports was partly shown to be exaggerated, and partly explained as compatible with the shared character of the intuition, if we understand it, not as an active tendency to believe, but as a disposition to undergo this tendency, activated when one has the right background concepts and engages in sufficient reflection.

The objection can also appeal to experimental philosophy, the subdiscipline that studies the intuitions of 'ordinary' people—non-philosophers—on philosophical topics.[23] Some studies suggest that ordinary people do not deploy a concept of phenomenal consciousness: they do not group paradigmatic phenomenal experiences together and accept, for instance, that robots can experience red while only humans can experience pain.[24] If people lack a concept of phenomenal consciousness, how can it appear to exist to them, or appear to be mysterious?

[23] Knobe and Nichols (2017).
[24] Sytsma and Machery (2010); Sytsma and Ozdemir (2019).

Various answers are available. Ordinary people could have a concept of phenomenal consciousness that does not guide these particular categorizations. They could read terms like "experience" or "consciousness" as only sometimes referring to phenomenal states.[25] They could treat various phenomenal experiences differently.[26] Finally, and this is the most plausible answer, they could have individual phenomenal concepts, each characterizing its referent as mysterious, but often fail to form the general concept.

Some studies directly investigate mystery intuitions, yielding contrasted results. One study suggests that mystery intuitions feeding the Knowledge Argument—roughly, the intuition of undetectability—are widely shared.[27] A small study suggests that the intuition of revelation is also widely shared.[28] Other studies found the opposite about irreducibility intuitions: most subjects do not find phenomenal consciousness irreducible (epistemically or metaphysically),[29] nor zombies possible.[30]

These last studies have problems. First, the terms they use to probe subjects are polysemic,[31] and we cannot assume that subjects read them correctly. For instance, if consciousness is irreducible, zombies must be *metaphysically* but not necessarily *physically* possible. Consciousness must be constitutively or reductively unexplainable, but could still be explainable in terms of covering laws. So, on some readings of "possible" or "explainable," consciousness is explainable, and zombies are possible, even if consciousness is irreducible, metaphysically and epistemically. However, it is plausible that ordinary people spontaneously use these readings: e.g., in ordinary contexts, "possible" means *physically* possible, not metaphysically possible.

Second, these studies probe mystery *beliefs*, not mystery *intuitions*: they ask subjects their opinions about the issues.[32] It is highly probable that at least *some* subjects have irreducibility intuitions but lack irreducibility beliefs, since it is the case with many philosophers of mind. The studies do

[25] M. Liu (2022).
[26] Chalmers (2020c, 239–40).
[27] Gregory et al. (2022).
[28] M. Liu (2021, 1500–1501).
[29] Díaz (2021).
[30] Peressini (2014).
[31] M. Liu (2022).
[32] This applies to Díaz (2021). Gregory et al. (2022) tried to probe specifically people's intuitions (and not just their opinions) by asking them whether they were intuitively inclined to give an answer different from the one they chose.

not tell us how numerous this group is and thus do not really show that irreducibility intuitions are scarce.[33]

In principle, experimental philosophy could confirm or disconfirm that mystery intuitions are shared. At this stage, the negative evidence it provides is weak and inconclusive; it also provides tentative positive evidence. None of this warrants abandoning the plausible presupposition that these intuitions are shared.

7.5. Do mystery intuitions result from our phenomenal concepts' characterizations?

Suppose mystery intuitions are shared and compelling. One objection is that they are not *conceptual*: they do not stem from the way phenomenal concepts characterize their referents. The objection can notably be raised about irreducibility intuitions.

Consider the Phenomenal Concept Strategy,[34] a popular defense of realist physicalism associated with type-B materialism.[35] It explains irreducibility intuitions by properties of our phenomenal concepts, but denies that these intuitions stem from how our concepts characterize their referents: it does not view irreducibility intuitions as *conceptual* in my sense.

In one version of the strategy, phenomenal concepts grasp phenomenal experiences (i.e., certain physical states) directly, without associated cognitive or descriptive content—for instance, in virtue of brute recognitional capacities.[36] This supposedly explains why zombies are *conceivable*: as phenomenal concepts lack cognitive content, no contradiction can be derived from the concept of a zombie. It also explains why zombies are *impossible*: if phenomenal experiences *are* certain physical states, physical duplicates are necessarily phenomenal duplicates.

If this view is correct, irreducibility intuitions are not conceptual in my sense. However, as argued in Chapter 1, it is implausible. Phenomenal concepts clearly provide a substantive grasp of phenomenal experiences, and positively characterize them. The view also struggles to explain the full range of irreducibility intuitions, which go beyond the mere negative

[33] For relevant remarks, see Chalmers (2020c, 241–42).
[34] For the expression, see Stoljar (2005).
[35] Chalmers (2002a).
[36] Carruthers (2000); Tye (2003); Levin (2007).

conceivability of zombies.[37] Phenomenal concepts' absent content might explain why phenomenal consciousness *does not seem* reducible to the physical, but it does not explain why *it does seem* irreducible to the physical. Distinguishing between *X not-seeming F* and *X seeming non-F* is key. If I use an empty verbal label ("A"), without any conception of A, I can wonder whether A is physical: the idea that A is not physical yields no contradiction. This happens even if "A," because of appropriate causal relations, refers to something physical. Still, it is not the case that A seems non-physical: I simply have no intuition about A's physical or non-physical nature. A is physical and does not seem physical, but it also does not seem non-physical. If phenomenal concepts' only peculiarity is their lack of cognitive content, they are comparable to the label "A," and they do not fully explain our irreducibility intuitions.[38]

In another version of the Phenomenal Concept Strategy, phenomenal concepts do not simply lack cognitive content. They also have a unique use/mention feature. Using phenomenal concepts, we always instantiate the phenomenal property (identical to a specific physical property) we represent: we activate an experience of the type thought about, or at least a copy of it.

Wondering whether phenomenal properties are physical, we use two concepts referring to the same properties: phenomenal concepts, whose applications activate the phenomenal experience thought about (or a copy thereof), and physical concepts, whose applications do not normally activate it. For Papineau, this generates:

> an intuitive sense in which exercises of [physical concepts] "leave out" the experience at issue. They "leave out" [. . .] the technicolour phenomenology, in the sense that they don't activate or involve these experiences.[39]

We then commit what he calls the "Antipathetic Fallacy," projecting this actual phenomenological difference between our two concepts on the referents of the concepts, which we conclude are distinct. This is how Papineau,[40] as well as many others,[41] explains our intuition of irreducibility. In this view,

[37] Levine (2001); Papineau (2008).

[38] Levine (2007, 147). There are deep parallels between this line of reasoning and my criticism against sparse-illusion views' accounts of irreducibility intuitions in Chapter 4.

[39] Papineau (2002, 170).

[40] Papineau slightly changed his account across time, but maintained the same core idea (Papineau 1993b; 2007).

[41] Loar (1990; 1997); Tye (1999, 712–13); Balog (2012a).

features of our phenomenal concepts explain our intuition of irreducibility, but these concepts do not characterize their referents as irreducible: our irreducibility intuition is not conceptual in my sense.

This view is unsatisfying. First, it encounters numerous counterexamples.[42] Second, it implies that irreducibility intuitions rest on a fallacy. However, many competent thinkers report or endorse these intuitions, and it seems unreasonable and uncharitable to suppose that all repeatedly commit the fallacy even after it was pointed out.[43]

Third, the process generating irreducibility intuitions in us seems psychologically different from processes generating fallacies or cognitive illusions. It rather seems similar to processes typical of valid reasoning.[44] Indeed, when we face fallacies, we can normally understand, on careful reflection, why what fallaciously seemed true is false, and vice versa. Take the conjunction fallacy,[45] where people fallaciously judge that Linda, described as outspoken, bright, having majored in philosophy, and concerned with social justice, is more likely to be a feminist bank teller than a bank teller. We might all tend to commit this fallacy. Careful reflection does not suppress this tendency, but it allows us to clearly understand *why* Linda is, in fact, more likely to be a bank teller than a feminist bank teller. This seems true of fallacies in general: upon careful reflection, we can clearly understand why the correct answer is what it is. In contrast, irreducibility intuitions do not work like this. Even on careful reflection, we struggle to form a clear and positive conception of phenomenal consciousness as being nothing over and above physical processes. For instance, in Chapter 1, I cited thinkers like Broad, Nagel, or Levine, who found the view that the phenomenal reduces to the physical "plainly nonsensical," "unintelligible"—a position we "cannot understand." However, our difficulty in forming a positive and coherent conception of the phenomenal as purely physical is naturally explained if my view is correct, that is, if phenomenal concepts implicitly characterize their referents as irreducible. Irreducibility intuitions do not rest on fallacies. This other alternative account also fails.

[42] See Sundström (2008). Papineau answers these counterexamples (Papineau 2011), but other, more sophisticated counterexamples can be raised (Kammerer 2018b)—though see H. Liu (2020) for a discussion.
[43] Melnyk (2003); Chalmers (2018, 32).
[44] Kammerer (2019b).
[45] Tversky and Kahneman (1983); Fisk (2004).

7.6. Are phenomenal concepts' problematic characterizations ineliminable commitments?

Suppose phenomenal concepts implicitly characterize their referents as mysterious—obvious, irreducible, and undetectable: mystery intuitions are conceptual. Suppose also that the view that nothing has these mysterious features is attractive.

Does this make eliminativism about the phenomenal attractive? Only if one also thinks that the relevant characterizations are *ineliminable*, *non-negotiable* commitments of the concept. One objection is that they are not.

This objection is often raised in conversation. It was recently pushed by Eric Schwitzgebel,[46] with a helpful example. Arguably, in Early Modern Europe, people used a king concept characterizing kings as rulers by divine right—God's lieutenants. They also used a witch concept, characterizing witches as having magical powers. Many later came to think that no one rules by divine right, and no one has magical powers. However, they concluded that there are still kings, even if there are no witches. The characterization HAS MAGICAL POWERS was an ineliminable and non-negotiable commitment of the concept WITCH, while RULES BY DIVINE RIGHT was merely an eliminable and negotiable commitment of KING. People appropriately *revised* their king concept, but not their witch concept.

Suppose phenomenal concepts characterize their referents as mysterious. This characterization may correspond only to eliminable commitments of these concepts. If so, knowing that nothing satisfies these problematic characterizations does not make eliminativism about the phenomenal more attractive than eliminativism about kings for someone confident that no one rules by divine right. The *Argument from Mysteries* is defanged. The case for eliminativism is threatened.

Why think that phenomenal concepts' problematic characterizations are eliminable commitments? Schwitzgebel suggests that, to know whether a commitment is eliminable, we must look at what people would conclude upon learning that the corresponding characterization is not satisfied. Suppose we come to believe that nothing is appropriately mysterious. Would we conclude that phenomenal concepts do not apply, like we did for WITCH, or would we conclude that they still apply, like we did for KING? In the first case, the commitment is ineliminable. In the second, it is eliminable.

[46] In personal correspondence, as well as in a blog post (Schwitzgebel 2025).

However, arguably, upon learning that nothing real is, say, obvious, irreducible, or undetectable, many of us would conclude that phenomenal consciousness exists but is not as it seems, rather than admit that it does not exist. They would choose weak illusionism over strong illusionism. This supposedly shows that the problematic commitments of phenomenal concepts are eliminable. Eliminativism is not attractive. The position of choice is not strong illusionism, but a weak realist, weak illusionist, *revisionary* view of phenomenal consciousness.[47]

Before I answer this objection, two precautions are necessary. First, we must carefully distinguish *terms* and *concepts*. The objection does not rely on the idea that people, upon learning that nothing is appropriately mysterious, would change the meaning of the *term* "phenomenal consciousness" to make it express a different concept—e.g., a quasi-phenomenal concept rather than a phenomenal concept. This sort of consideration leads to a different objection, discussed below—the redefinition objection. The idea here is rather that people would use a revised version of the same concept: they would conclude that *phenomenal consciousness*, not just quasi-phenomenal consciousness under another name, really exists, but lacks mysterious features.

Second, the objection only makes sense if we assume that there is a genuine difference between *revising* and *replacing* concepts. This requires concepts to have individuative, ineliminable cores, distinct from their non-individuative, eliminable components. Changing the second leads to conceptual revision; changing the first leads to conceptual replacement. Those who deny that there is a difference between revising and replacing concepts, for instance, because they deny the existence of stable conceptual cores (say, for Quinean reasons), will not run this objection. Of course, they can make a related objection about *terms*—the redefinition objection, discussed below.

For now, assume, for the sake of the argument, that concepts have individuative ineliminable cores. The core of WITCH, for instance, includes the characterization HAS MAGICAL POWERS. These concepts' cores are distinct from their non-individuative eliminable commitments, like RULES BY DIVINE RIGHT for the concept KING. I do not presuppose a view of what makes a characterization a part of the individuative ineliminable core of a concept, but I accept the following criterion. Suppose a subject believes with full confidence that an object O fails to satisfy characterization S of concept

[47] For revisionism about phenomenal consciousness, see Lewis (1995); Chrisley and Sloman (2016); A. Clark et al. (2019).

C. If S is a core part of C, it can never be rational to think that C applies to O. If S is a non-core part of C, it can.

Come back to the objection. Grant that many people, upon learning that nothing is appropriately mysterious, would conclude that phenomenal consciousness exists, but lacks mysterious features. Does this show that phenomenal concepts' problematic characterizations are merely eliminable commitments?

No. To determine the eliminability of problematic characterizations, what matters is not what people *would* conclude upon learning that nothing is appropriately mysterious, but what they *should rationally conclude*. The two issues are related, but distinct: people's conclusions are shaped by many factors, and can depart from the rational response. This is obvious on reflection: otherwise, the mere fact that weak illusionism is more popular than strong illusionism among those who deny mysterious features would justify concluding that it is more rationally compelling.

A non-phenomenal example also makes this vivid. Imagine Bernadette, a fervent Catholic who first approaches theological subtleties. She is told that there is only one God, that the Father is God, and that Jesus is God. She believes all this with full confidence. At this point, rationally, she should conclude that *Jesus is the Father*. However, she does not. She is so strongly disposed to believe that Jesus is not the Father that, to accommodate this last belief, she accepts a Trinitarian view which, by the Church's own admission, is unintelligible—a *mysterium fidei*. Bernadette's conclusion departs from the rational response.

Come back to the consciousness case. Suppose that most people would indeed react to the denial of mysterious features by embracing weak illusionism, not strong illusionism. Different diagnoses are possible. One diagnosis, on which the objection relies, is that the characterizations-as-mysterious are eliminable commitments. We are in a king-like case: the concept of phenomenal consciousness can be stripped of its implicit problematic commitments, without changing its core.

A rival diagnosis, which I think is correct, is that the characterizations-as-mysterious are *ineliminable*. We are in a witch-like case: the concept of phenomenal consciousness cannot be stripped of its implicit problematic commitments, which belong to its core. However, people are so strongly and robustly disposed to believe that they are phenomenally conscious that they tend to accommodate this belief, even if this means departing from the rational response: we are in a Trinity-like case. As a result, many people embrace views that exhibit a degree of implicit incoherence and unintelligibility: views that assert the existence of phenomenal consciousness, and thus implicitly

assert that something real is obvious, irreducible, and undetectable, while explicitly denying that anything is obvious, irreducible, or undetectable.

Two considerations plead for the second diagnosis. First, we have independent reasons to believe that some of its key posits are real. I argued earlier that we indeed have a very strong disposition to believe that we are phenomenally conscious. The Trinity example shows that strong dispositions to believe can make us diverge from the rational response and adopt incoherent views. These facts do not prove that the second diagnosis is true, but they make it prima facie plausible.

Second, the second interpretation makes specific confirmed predictions. If it is correct, many who conclude that phenomenal consciousness exists but lacks mysterious features should feel persistent epistemic discomfort. They should find their own view somewhat unsatisfying, if not incoherent. They should be tempted to admit that the truth of their view remains an incomprehensible mystery. The first interpretation does not make the same predictions. These predictions are confirmed: as stressed in Chapter 1, many physicalists (or thinkers attracted to physicalism) admit that physicalism "seems unintelligible," "is a position we cannot understand," and is something we are stopped from "really believing."[48] *Mysterianist* versions of materialism have been found tempting since the mid-19th century (Griesinger, Huxley, du Bois-Reymond) and remain so to this day.

Incidentally, if the first interpretation were correct and our concept of phenomenal consciousness, like KING, only had eliminable problematic commitments, I think that we would probably *already* have disentangled the core unproblematic concept and its eliminable problematic commitments, the way we disentangled KING from GOD'S LIEUTENANT. Plausibly, we did not do so, precisely because our concept of phenomenal consciousness was *already* the result of this kind of disentanglement: it is the implicitly problematic residue remaining after we unraveled other characterizations attached to our multi-criterial ordinary concept of consciousness. We took this general concept CONSCIOUSNESS, chipped away at all the unproblematic characterizations (related to what we call access-consciousness, self-consciousness, etc.), and were left, precisely, with our concept of phenomenal consciousness.

If I am right, phenomenal concepts' problematic characterizations are ineliminable commitments. Supposing that nothing in reality has the right

[48] Nagel (1974, 446); Papineau (2002, 95); Levine (2007, 148).

mysterious features, the rational response is to admit that there are no phenomenal experiences. Eliminativism remains attractive.

The idea that problematic characterizations are ineliminable commitments of phenomenal concepts follows from Meta-Cartesianism. The view entails that we represent phenomenal experiences as primitive direct evidence, and essentially think of them *as obvious*. There is no substantive and specific characterization carried by our phenomenal concepts that could be disentangled from this problematic characterization as obvious. Our understanding of the phenomenal as irreducible and undetectable also quickly follows, given other shared concepts.

Suppose you agree with all this. You might respond that, still, we should not deny phenomenal experiences. Biting the bullet and admitting, with strong realists, that experiences really have mysterious features is preferable to denying phenomenal experiences. Endorsing a realist physicalist view about phenomenal consciousness, which for us remains as mysterious as the Trinity, is preferable to denying phenomenal experiences. Any realist view is preferable to denying phenomenal experiences. Why? Because the existence of phenomenal experiences is *obvious*. It is more certain than any opposing consideration. This is the Moorean argument against illusionism, to which I now turn.

7.7. The Moorean argument against illusionism

Suppose that illusionism is coherent and supported by strong arguments. All objections have been answered. The last line of defense of realists is the idea that we can still confidently rule illusionism out, because it is obviously false. This is the Moorean argument against illusionism.[49]

In conversation, this objection takes the form of an incredulous stare or a shrug. An early philosophical articulation is due to James Cornman, who introduced the label "eliminative materialism." Cornman thought the best response to eliminativism was this:

> Those terms that denote the confirmation basis of all knowledge, i.e. the phenomenal terms that denote the data known by acquaintance, describe what we can be most certain there is.[50]

[49] For a first version of my treatment of this objection, see (Kammerer 2022c).
[50] Cornman (1968a, 62).

For Cornman, we can reject eliminative materialism "no matter what the future of science," since it is manifestly false, and negates what we can be most certain there is. He reiterated this point elsewhere:

> It may be at least plausible to claim physical objects are quite different from what most of us believe, because they are best described by the theoretical terms of science, and it may be plausible to claim that our sensory phenomena have some (e.g., neurophysiological) features we do not experience. But it is most implausible to claim that a man's sensory phenomena have nothing like the features he experiences them to have, with the consequence that he has no special epistemological status even regarding those features he believes his sensory phenomena have.[51]

A version of the objection appears in the work of Bryan Frances[52]:

> Even if all the philosophers of mind and cognitive scientists were screaming at my door "No one has ever had a painful feeling! There are no feelings! No one has ever been in pain!" I would still know that I have had loads of painful feelings. My "access" to my painful feeling is so reliable or direct that I can epistemically neutralize, without even thinking about it, any hypothesis that says I have no feelings. I might be wrong about the location of the feeling, but there's no way in hell I'm wrong that I have a painful feeling.[53]

Finally, Chalmers labeled the objection the "Moorean argument against illusionism," as it resembles Moore's proof of the external world.[54] Chalmers's version, on which I focus, goes like this[55]:

> **Premise 1**: People sometimes feel pain.
> **Premise 2**: If illusionism[56] is true, no one feels pain.
> **Conclusion**: Illusionism is false.

So formulated, Chalmers's second premise is debatable. Illusionists grant that we feel pain in a *functional* sense: we enter states that cause screaming,

[51] Cornman (1968b, 35).
[52] See also Strawson (1994, 101).
[53] Frances (2008, 588).
[54] Moore (1939).
[55] Chalmers (2018, 44–49).
[56] Chalmers writes "strong illusionism."

avoidance behavior, trauma, etc. They also grant that we feel pain in a 'normative' sense: we enter states that are awful for us and make us deserving of compassion and care. (More on this in Chapter 9.) If, as I claimed, taking inspiration from Papineau, our ordinary talk of "consciousness" (tout court) typically expresses a multi-criterial concept, plausibly satisfied as long as at least some criteria (e.g., functional ones) are satisfied even if others (e.g., phenomenal ones) are not, then illusionists can also truthfully say that we feel conscious pain. (More on this in the next section.)

Still, it is true that illusionists deny that we feel *phenomenal* pain. On this specific reading of "feeling pain," they accept Chalmers's second premise. From now on, I assume that this is what is meant here.

The key Moorean premise is the first premise: people sometimes feel phenomenal pain. For Chalmers, it is an "introspectively obvious datum." To avoid confusion with what I called earlier "introspective obviousness," I will say that, for Chalmers and other proponents of the objection, the premise is *manifest*: it has very strong pre-theoretical justification.

Moorean arguments, in general, are direct defenses of manifest commonsensical claims against philosophical arguments to the contrary.[57] They go back to Moore's "proof of an external world." They are often accepted by contemporary philosophers, who still debate why and how they work.[58] For the sake of the argument, I grant that Moorean arguments sometimes work. Some commonsensical premises of Moorean arguments are manifest; their strong justification is transmitted to their conclusion negating the conclusion of a philosophical argument, and the result is dialectically effective.[59]

Still, the received view is that Moorean arguments only counter purely philosophical arguments, but not scientific arguments. As William Lycan puts it:

> Common-sense beliefs can be corrected, even trashed entirely, by careful empirical investigation and scientific theorizing [. . .] Common sense must yield to evidence [. . .] but it need not yield to bare metaphysical

[57] See Lewis (1973b, 88); Fine (2002, 2); Gupta (2006, 178); Kelly (2005). References found in Rinard (2013).

[58] Some think Moorean arguments are a matter of preferring the most plausible premises (Lycan 2001, 38–39). Others think they rely on the idea that we should pay attention to judgments about particular cases as opposed to general principles (Kelly 2005), or a general principle of theoretical conservatism (Harman 2003). For a presentation, see Rinard (2013, 198–211). I use the first—and least committal—interpretation.

[59] See Pryor (2004, 350, 359) for various ways of rejecting Moorean arguments.

> pronouncement [. . .] No *purely philosophical* premise can ever (legitimately) have as strong a claim to our allegiance as can a humble commonsense proposition such as Moore's [. . .]. Science can correct common sense; metaphysics and philosophical "intuition" can only throw spitballs.[60]

For example, we think that modern physics—special relativity—shows that no two distinct events can happen at the same time independently of a reference point. There is no absolute simultaneity, even if the claim that there is absolute simultaneity is commonsensical and manifest.

If so, the Moorean argument against illusionism is in a bad position. Indeed, the illusionist arguments are not purely philosophical arguments. They are scientific-cum-philosophical arguments, featuring crucial premises justified by science. For instance, the *Argument from Mysteries* appeals to *Explanation*: the thesis that realism does not account for our non-phenomenal data better than eliminativism. This premise is notably supported by reflecting on the commitments of our best scientific accounts of physical data. It is also justified by reflecting on the causal closure of the physical, itself made plausible by the success of natural science. The *Argument from Irrelevance* appeals to a *Local Explanation*, a special case of *Explanation*.

Illusionist arguments are not purely philosophical. They are philosophical interpretations of scientific successes. This entails that they are not *purely scientific* either: they require philosophical premises. However, it is also the case of those scientific arguments we think justify rejecting common sense, like the argument against absolute simultaneity from special relativity. As Susanna Rinard convincingly argues, this argument requires preferring special relativity over empirically equivalent neo-Lorentzian alternatives, a preference justified by philosophical principles favoring simpler views.[61]

Rinard herself goes as far as claiming that, if philosophical interpretations of science can overturn common sense, so does philosophy alone, since scientific-cum-philosophical arguments can only be as strong as their weakest premises—arguably, the philosophical one. I do not follow her there. I think that scientific-cum-philosophical arguments are able to overturn common sense, while purely philosophical arguments might not be.

[60] Lycan (2001, 40–41). Rinard (2013) notes that Gupta (2006, 178) and Kelly (2008) share this view.

[61] Rinard (2013, 189–98).

Something dialectically peculiar, changing the assessment of the whole argument, might occur when arguments contradicting common sense have at least one premise that is appropriately crucial, antecedently surprising, *and* science-based.

At any rate, we see that the Moorean argument against illusionism is in bad company: it stands with putative Moorean arguments for absolute simultaneity, not with the Moorean argument for the existence of the external world. By this, I do not claim that illusionist arguments are as strong as the argument against absolute simultaneity, only that they are of the same general type.

In response, proponents of the anti-illusionist Moorean argument would stress that their argument is different from standard Moorean arguments. Its key premise is *more* manifest than the Moorean premises at play in standard Moorean arguments. This makes the argument able to counter even science-based arguments. Arguably, this is the view of Cornman, Frances, and Chalmers.

Call a commonsensical premise which is more manifest than standard manifest claims, and so manifest that it can counter even science-based arguments, a "super-manifest" premise. Call the arguments powered by such premises "super-Moorean" arguments. For us to be able to rule illusionism out confidently, the claim that phenomenal pain exists must be super-manifest, not just manifest. The Moorean argument against illusionism must be super-Moorean.

Is the claim that phenomenal pain exists really super-manifest? The question might seem inappropriate. Manifest claims do not require further justification. However, when asking this question, we do not ask to *justify* the claim that phenomenal pain exists. Either this claim is super-manifest, or it is not. If it is, the anti-illusionist argument is successful, even without further justification. However, the question bears on the *second-order claim* that the claim really is super-manifest. Asking this second-order question about putative super-Moorean arguments is reasonable. Claims thought to be super-manifest could in fact be merely manifest, or not manifest at all. Past failures to rebuke science-based arguments on the basis of putative super-manifest claims show that we should ask the question. Think of Bergson in the 1920s,[62] attacking the relativistic claim that there is no absolute simultaneity by appealing to an intimate epistemic relation to time—what Bergson

[62] Bergson ([1922] 1999).

called "duration" (*durée*). Arguably, he thought that the existence of absolute simultaneity was sufficiently manifest to rebuke scientific arguments, but we now think he was wrong. We should be cautious when facing alleged super-manifest claims.

Is the existence of phenomenal experiences super-manifest? Some think so. They think that we have a special epistemic relation to phenomenal experiences—a uniquely strong pre-theoretical justification to believe they exist. This epistemic relation can be cashed out in terms of introspective obviousness, acquaintance, etc. Call it "acquaintance" here.

Illusionists deny acquaintance. They can build arguments against its existence. First, an *Argument from Mysteries**, stressing that acquaintance is mysterious, and that we can explain physical data as well if we deny acquaintance as if we accept it—the conclusion being that, if we have good reason to reject direct data about acquaintance, we should deny acquaintance. They can then build an *Argument from Irrelevance**, premised on the claim that the physical data regarding our judgments about acquaintance is explained as well if we deny acquaintance as if we accept it, which ultimately gives a reason to reject direct data about acquaintance.

I do not detail these arguments here: I assume that, if we have good science-based arguments against phenomenal consciousness, we have good science-based arguments against acquaintance. In fact, given Meta-Cartesianism, anti-acquaintance arguments and anti-phenomenality arguments *should* overlap, since we represent phenomenal experiences as introspectively obvious entities—entities with which we are acquainted.

If proponents of the anti-illusionist Moorean argument admit that we are not acquainted with phenomenal experiences and that their existence is not super-manifest, they concede that their first argument is unsuccessful. They must stand their ground. To answer the illusionist anti-acquaintance argument, they can provide a Moorean or a non-Moorean argument.

The first option is attractive, but ultimately hopeless. A Moorean argument for acquaintance must be *super-Moorean*, since anti-acquaintance arguments are science-based. Illusionists can then build a second-order science-based argument against acquaintance being super-manifest—against our acquaintance with acquaintance. The Moorean would have to produce a third-order super-Moorean argument for acquaintance with acquaintance, requiring us to be acquainted with this acquaintance with acquaintance, etc. The same dialectic can be reiterated a number of times. At some point, however, the Moorean cannot pull off the same move: there

must be some level *n* such that it is not manifest at all that we have the right epistemic relation at *n*−1, because it is simply implausible that the existence of epistemic relations involving arbitrarily high levels of complexity is pre-theoretically justified. At some point (be it the fifth-order, the tenth-order, etc.), people simply cannot have strong pre-theoretical intuitions anymore. But, if at level *n*, the Moorean grants that it is *not* super-manifest that they have a unique epistemic relation at *n*−1, they cannot counter the science-based argument against the existence of this relation at *n*−1. They must concede that it does not hold, and grant that the science-based argument against the existence of this relation at *n*−2 succeeds, etc. Ultimately, they concede that the existence of phenomenal experiences is not super-manifest. The Moorean argument against illusionism fails.

The second option is to give a non-Moorean defense of acquaintance (or acquaintance with acquaintance, etc.), which does not merely appeal to manifest claims, but mobilizes scientific or philosophical views. However, if a non-Moorean argument is needed to defend the idea that the existence of phenomenal experiences is super-manifest in the way required for the anti-illusionist Moorean argument to succeed, this non-Moorean argument is doing the core dialectical work here. We must weigh it against illusionist arguments. This weighing is what really decides the issue. In this case, the anti-illusionist Moorean argument might be successful, but it is, in itself, dialectically irrelevant. Moreover, although I will not argue for this, it is plausible that views required to power this non-Moorean defense of acquaintance imply the falsity of illusionism, rendering this defense question-begging.

Let us take stock. Some contend that illusionism can be ruled out because the existence of phenomenal experiences is manifest. Their argument only succeeds if it is not just manifest, but *super-manifest.* But is it? Illusionists can build a science-based argument that it is not. Anti-illusionists can give Moorean reasons to think it is. The dialectic moves to a higher level, and illusionists win the debate as higher-order complexity makes us leave the shores of common sense. If anti-illusionists defend the super-manifest character of phenomenal experiences' existence in a non-Moorean manner, this defense does the argumentative heavy lifting, rendering the Moorean argument dialectically irrelevant.

Giving up on the Moorean argument does not mean accepting illusionism. It does not mean abandoning the idea that phenomenal experiences are "ordinarily" manifest. It simply means admitting that illusionism cannot be ruled out conclusively simply because the existence of phenomenal

experiences is manifest. Illusionism is counterintuitive, but this is not enough to discard it conclusively. The existence of phenomenal experiences may be manifest, and we may have strong pre-theoretical reasons to believe they exist, but these reasons are not strong enough to spare us the duty to examine and weigh the arguments.

Finally, Meta-Cartesianism explains why the existence of phenomenal experiences seems super-manifest. We represent them as introspectively obvious—as such that we have, about them, infallible direct evidence. We thus naturally tend to think of their existence as infallibly justified in a pre-theoretical manner—super-manifest. This explains why the Moorean objection is so tempting.

7.8. The "Now what?" question and the redefinition objection

A. The "Now what?" question

Suppose you accept strong illusionism. What to do with beliefs and discourse about consciousness in general, and about phenomenal consciousness in particular? This is the "Now what?" question for illusionism. The name comes from meta-ethics, where it refers to the problem of knowing what to do with moral beliefs and discourse once we accept moral error theory.[63]

First, consider *ordinary* beliefs and discourse about consciousness in general—consciousness *tout court*. Those are widespread. We often think and talk about creatures and states being conscious. Following Papineau, I think that our ordinary concept of consciousness is a complex multi-criterial concept; one of these criteria corresponds to phenomenal features, others to functional features.

If I am right, once we embrace illusionism, what is most natural is for us to keep believing and saying that consciousness exists, that we are conscious creatures, and that some of our thoughts and perceptions are conscious. Indeed, the right functional features, arguably warranting applications of the concepts, are probably instantiated even if illusionism is true. Some of these functional features might correspond to access-consciousness or self-consciousness; others to what illusionists call "quasi-phenomenal consciousness"—i.e., the set

[63] Joyce (2005); Lutz (2014); see Jaquet (2019) for a dissenting take on the possible answers to the question. My discussion of the "Now what?" question for illusionism only loosely resemble the discussions in meta-ethics, since constraints on answers differ in the two domains.

of mental states mischaracterized as phenomenal in introspection. This does not matter much here. The one change brought by illusionism is that, once we accept the view, we should think and say that, while consciousness *tout court* is real, it lacks features it seems to have. It notably lacks phenomenal features. Embracing strong illusionism about phenomenal consciousness, we should accept weak illusionism about consciousness tout court.

Second, consider *scholarly* beliefs and discourse about consciousness tout court. Arguably, those are not as widespread. Consciousness researchers often try to disentangle the components of our ordinary concept. They usually avoid statements about consciousness tout court. Strong illusionism simply gives one more reason to be thorough in this respect, to avoid confusing existing functional features with non-existent phenomenal ones. More than ever, scientists and philosophers should avoid talking of *consciousness* without specifications. I come back to this point in the next chapter, dedicated to consciousness science.

Now, consider *ordinary* beliefs and discourse specifically about *phenomenal consciousness*. Arguably, these are rare. Ordinary thought and talk rely on our multi-criterial general concept, and no ordinary term unambiguously refers to phenomenal consciousness: "phenomenal," "qualia," "what it's like," etc., are technical terms, or ordinary terms used technically. Still, sometimes, in non-scholarly contexts, we think and talk, without dedicated words, about phenomenal experiences specifically. Think of the day you ate your first dragon fruit and described your experience. Think of Marcel Proust describing a complex stream of feelings and memories he had after tripping on an uneven paving stone. What should illusionists do with the corresponding beliefs and utterances?

It helps to adapt the options open to the ethical error theorist facing the "Now what?" question. Like error theorists, illusionists can choose between:

(1) An *abolitionist* stance, where they abandon phenomenal beliefs and corresponding utterances.
(2) A *conservationist* stance, where they maintain these beliefs and corresponding utterances, while theoretically knowing that they are all false.
(3) A *fictionalist* stance, where they replace phenomenal beliefs with pretenses, and treat phenomenal utterances as fictional statements.
(4) A *substitutionist* stance, where they replace phenomenal beliefs with beliefs about something else (e.g., quasi-phenomenal states), and

reinterpret what looks like phenomenal utterances as utterances expressing these beliefs.

Of course, there are also hybrid stances. I will not argue for it at length, but I think that, if psychologically possible, the substitutionist stance is preferable in ordinary contexts. It leads to truthful and coherent beliefs and utterances (vs. conservationism and fictionalism). It does justice to the rich and correct informational content nested in our (false) phenomenal beliefs while minimizing cumbersome linguistic revisions (vs. abolitionism).

B. The redefinition objection

Finally, we are left with the case of s*cholarly* beliefs and utterances about *phenomenal* consciousness, held in philosophy and science. These are relatively widespread: thousands of articles and books discuss phenomenal consciousness. About these, a certain version of substitutionism is superficially attractive and leads, as I will show, to the redefinition objection, which is a verbal objection to illusionism.

Indeed, here is a way the substitution could go. Suppose you become convinced by illusionism. You could replace most of your old phenomenal beliefs with resembling quasi-phenomenal beliefs. You would only have to drop those that are clearly false of quasi-phenomenal consciousness, but this would still leave a lot of beliefs—e.g., those about who has which phenomenal experiences. You could also reinterpret most utterances about the phenomenal as bearing on the quasi-phenomenal. You would then redefine the expression "phenomenal consciousness" so that it expresses a concept of quasi-phenomenal consciousness. You could keep on expressing your old concept of phenomenal consciousness with an expression like "phenomenal consciousness as it seems to be."

Having done this, you would have the same substantive beliefs as strong illusionists, but talk like a weak illusionist: e.g., saying things like "phenomenal consciousness exists, but is not as it seems to be." You would adopt *verbal weak illusionism*. Verbal weak illusionism is a revisionary approach to phenomenal consciousness, except that what is revised here is not the concept, but the terms.

Verbal weak illusionism is substantively identical to strong illusionism. There can be no epistemic reason to choose one over the other. Still, we can

wonder whether we *should* operate this redefinition. If we have good non-epistemic reasons to prefer verbal weak illusionism over strong illusionism, we gain a non-epistemic objection to strong illusionism.

The possibility of verbal weak illusionism, substantively equivalent to strong illusionism, has long been acknowledged. Pereboom's view is a form of verbal weak illusionism, and Pereboom himself sees the difference between his view and strong illusionism as "merely verbal and non-substantive."[64] Christopher D. Brown and David Papineau think that there is no substantive difference between strong illusionism and realist a posteriori physicalism,[65] which implies that verbal weak illusionism is possible and is no other than a posteriori physicalism rightly interpreted. Dennett explicitly endorsed strong illusionism,[66] but viewed the choice between strong and weak illusionist formulations of the same substantive view as a matter of "diplomatic policy," not substantive disagreement.[67]

Michael Graziano and Nick Humphrey went further, putting forward versions of verbal weak illusionism, which they claimed is superior to strong illusionism.[68] They endorsed a version of the redefinition objection. From their perspective, why prefer verbal weak illusionism to strong illusionism? Roughly, because strong illusionism sounds radical, if not crazy: someone who agrees with its substance might wish for more palatable formulations. Verbal weak illusionism sounds less radical. Therefore, it is rhetorically and strategically superior. It maximizes the view's acceptability without substantive loss.

For instance, Graziano agrees that his Attention Schema Theory is "a kind of illusionism," and concedes that the "illusionist approach" is "essentially correct." Still, he warns against illusionist formulations:

> In my experience [. . .] calling consciousness an illusion is the kiss of death for a theory.[69]

This led him to something like verbal weak illusionism. Nicholas Humphrey made a similar point, noting that:

[64] Pereboom (2019, 188).
[65] C. D. Brown and Papineau (2024). The claim was anticipated by Quine (1960, 264–65).
[66] Dennett (2016; 2019).
[67] Dennett (1991, 459).
[68] Humphrey (2016); Graziano (2019a).
[69] Graziano (2019a, chap. 7). Graziano explicitly targets illusionist formulations denying *consciousness*, but I use his remark to discuss those denying *phenomenal consciousness*. It is not clear to me that illusionist formulations have unambiguously targeted consciousness tout court.

> illusionism cannot but feed ordinary people's fears that we scientists/philosophers *want to take consciousness away from them*.[70]

He put forward a view he called "phenomenal surrealism," which I read here as a form of verbal weak illusionism.[71]

So, the objection goes, verbal weak illusionism is preferable to strong illusionism because it avoids unnecessarily antagonizing statements, which makes it more acceptable. However, I think that, on reflection, this does not render verbal weak illusionism preferable—quite the contrary.

Indeed, what makes strong illusionism so hard to accept? Probably not a mere attachment to terms like "phenomenal consciousness" or "qualia," since these are fairly recent technical terms. More plausibly, strong illusionism is hard to accept because of its substantive content—because it denies the existence of something that seems obvious. However, this substantive content is upheld by verbal weak illusionism: strong illusionism and verbal weak illusionism are substantively identical. Both deny the existence of the same apparently obvious entity.

This suggests that verbal weak illusionism is only more acceptable than strong illusionism because it *seems* to have a different content: because it somehow disguises or obfuscates its substantive content. If so, verbal weak illusionism maximizes acceptability at the cost of *and in virtue of* a loss of clarity and intelligibility. It is more acceptable only because it is more likely to be misunderstood: people are more likely to miss that it denies something manifest.

This gives a powerful reason to prefer strong illusionism over verbal weak illusionism in scholarly contexts. Plausibly, in these contexts, clarity and intelligibility should be strongly prioritized over acceptability. Strong illusionism is hard to swallow, but it *should be* hard to swallow if correctly understood.[72] In scholarly contexts, a core dose of abolitionism is unavoidable at the level of utterances. We should unambiguously deny the existence of phenomenal consciousness. This is compatible with a dose of partial *doxastic* substitutionism: some of our old phenomenal beliefs could be replaced with quasi-phenomenal equivalents. This might notably be the case of beliefs concerning *which* phenomenal experiences we have, or about their causal

[70] Humphrey (2016, 121–22).
[71] Though see the discussion in my *Coda*.
[72] For a related point, see Frankish (2023, 10).

roles. In the next three chapters, I examine the implications of illusionism for consciousness science, ethics, and epistemology. These chapters can be read as examining whether various substantive beliefs involving phenomenal experiences held in these domains can or should be substituted with similar beliefs about, among others, quasi-phenomenal states.

Limiting this potential partial substitution to the doxastic domain is crucial. At the linguistic level, no substitution should occur. We should always make it as clear as possible that we deny phenomenal consciousness. Note that a key motivation for linguistic substitutionism in ordinary contexts is the minimization of cumbersome linguistic revisions. This does not have much import in scholarly contexts. Linguistic revisions focusing on technical terms ("phenomenal consciousness," "qualia," etc.) used in scholarly contexts are rarer and more concentrated, which makes them only moderately costly: it is worth making them if they bring substantive clarity gains. This is in line with historical practice. In past scientific and philosophical revolutions, we typically treated ordinary terms and terms of art differently. We usually *redefined* terms that originated in ordinary contexts to continue using them truthfully. Scientists kept talking about colors, time, solidity, and matter, even after physics taught us these things are very different from what we intuit them to be. To do so, they gave new meanings to these terms. In contrast, terms of art originating in scholarly contexts often (though not always—think of "atom") received a different treatment. We *denied* the existence of aether or phlogiston, rather than redefining these terms to make them express the concept of something real, as would have been possible in principle.

My preference for strong illusionism over verbal weak illusionism is non-epistemic and context-dependent. In *some* situations, we should opt for verbal weak illusionism, even as scholars. Imagine a dystopic future where a realist neo-Inquisition chases strong illusionists to execute them. In this eventuality, we should certainly obfuscate our views and become verbal weak illusionists, like the 17th-century *libertins érudits* who disguised their atheism by writing "in a manner and style [verging on] the secretive."[73] Alternatively, suppose strong illusionism becomes the received view. The expression "phenomenal consciousness" could become a faster and innocuous way to talk about quasi-phenomenal consciousness without confusion, since

[73] Fisher (2024).

everyone would admit that the phenomenal does not exist. This could make verbal weak illusionism preferable to strong illusionism.

7.9. Miscellaneous worries

Finally, illusionism raises other worries worth addressing. First, some people reject the view at the aesthetic or emotional level. They find it ugly or sad.

In one version of the worry, illusionism's sadness or ugliness stems from its normative consequences, as it supposedly describes a world devoid of well-being or moral status. This concern is addressed in Chapter 9. Another version of this worry concerns what illusionism linguistically implies. One could think it forces us to reject all phenomenal utterances, which seems regrettable. Do we have to throw away countless poems and novels about feelings, including seven volumes by Proust alone? When my wife tells me she experiences deep feelings of love in my presence, do I have to tell her she is wrong?

Luckily, illusionism has no such linguistic implications. Various responses to the "Now what?" question are open. The kind of substitutionism I recommend in ordinary contexts allows us to understand what is correct and insightful in Proust's description, and how my wife's declaration can be treated as true. Even if substitutionism turns out problematic or psychologically impossible, nothing forces illusionists to adopt the radical abolitionist stance that would have the regrettable consequences just mentioned. Other options, such as fictionalism or conservationism, remain viable.

Despite this, people may still have negative emotional or aesthetic reactions to illusionism. However, these reactions should not count against the theory. Many found heliocentrism sad—we are not at the center! Others found evolution theory ugly—we are apes' cousins! It is now evident that these reactions should not have counted against the views. The only beauty that plausibly matters for theory-choice is related to theoretical elegance, but there is little reason to find illusionism theoretically inelegant.

Still, if people continue to find illusionism ugly and sad, it is a strategic weakness of the view, as it would remain less likely to be accepted, even if it is true. This sort of weakness is more problematic in philosophy than in the sciences: as the pragmatic advantages of adopting correct theories are less straightforward in philosophy, pleasant philosophical errors are less likely to be weeded out than their scientific counterparts.

The issue might fix itself: as we become familiar with illusionism, its apparent sadness or ugliness might vanish. People may first dislike illusionism because it seems to deprive them of something, but as they get used to the view, they could realize that it only denies something they never had, and stop caring. Alternatively, illusionists could try to change how the view is perceived. The beauty and joy we associate with philosophical ideas are not intrinsic to them. They stem from our reactions, shaped by the associations and metaphors we deploy to grasp them. To many, an illusionist world seems dead and gray, in contrast with the rich, rainbow-like, and shimmering world of phenomenal realists. However, we could frame the opposition differently.

Realist views arguably have inaesthetic features. Focusing on the distribution of consciousness, panpsychism seems tacky, like the palace of a nouveau riche whose architects overused gold. Non-panpsychist realism appears clumsy, with its arbitrary threshold where two creatures with virtually similar cognitive capacities radically differ phenomenally—one being phenomenally conscious, while the other is not. Focusing on the metaphysics of consciousness, realist non-physicalism asks us to believe in fairies. Realist physicalism seems an insult to reason, forcing us to accept metaphysical claims we do not comprehend, like an incompetent teacher unable to explain a topic, and who resorts to banging on the table: "consciousness just is such and such material processes, *period*—too bad if this makes zero sense to you." In contrast, illusionism avoids distributional extravagance (vs. panpsychism) and unsightly threshold effects (vs. non-panpsychism). It is ontologically serious (vs. non-physicalism) and, while counterintuitive, avoids unintelligible metaphysical pronouncements (vs. realist physicalism).

There is also a strange heroic beauty in the illusionist narrative. We are the victims of a strong and pervasive illusion. Reality is radically different from what it seems to be. What we intuitively think of as our most certain and direct apprehension of reality, phenomenal introspection, is just another occasion for misrepresentation. Empirical science and philosophical reflection make us understand, against all odds, that we live among illusions, and we can bootstrap ourselves to a more accurate but counterintuitive view of reality, leaving the phenomenal house of mirrors. This story is exciting. Of course, it is neither correct nor incorrect in any deep sense, but it is available, and readers who reject illusionism mainly because it feels ugly and sad should consider it.

Finally, worries arise about illusionists themselves. Their sincerity is sometimes doubted. This philosopher defends illusionism in the seminar

room, but complains about his shoulder pain at the coffee machine: Is he even sincere? The previous discussion of the "Now what?" question answers this worry. Someone can coherently endorse illusionism, and nevertheless make sincere utterances that sound like phenomenal utterances (but do not express phenomenal beliefs), provided they employ a substitutional or fictionalist stance. Someone can sincerely, but not entirely coherently, be an illusionist and have phenomenal beliefs, taking a conservationist stance. We do not have to suppose that insincerity has to happen here.

Illusionists' epistemic virtues are also criticized. Some claimed that illusionists suffer from a naïve and "irrational attraction to scientism,"[74] or that they "find it exciting to hold views that seem preposterously contrary to common sense,"[75] as if heroic stories, or the desire *d'épater le bourgeois*, were really the main reasons why some people endorse the view. In response, the move leading to illusionism is nothing faddish. Illusionism is a relatively recent view, but the fundamental idea that things are not what they seem, and that we live in a world of appearances, is as old as philosophy. Illusionism, too, is a footnote to Plato. As for the view that we should crucially rely on science to decide what exists, it is one of the most widespread approaches to metaphysics nowadays. It does not require blind faith in science, and it is perfectly compatible with the view that our scientific theories are incomplete, and at best approximately correct.

* * *

This chapter addressed objections and worries against illusionism in general. I now move to the last part of the book, examining illusionism's implications for consciousness science (Chapter 8), ethics (Chapter 9), and epistemology (Chapter 10).

[74] Goff (2016, 96); see also Balog (2016).
[75] Strawson (2018).

8

The Science of Sirens

Illusionism and Consciousness Science

Mais où est le regard lumineux des sirènes
Il trompa les marins qu'aimaient ces oiseaux-là . . .

—Apollinaire

What implications, if any, does illusionism have for the science of consciousness? This is the question this chapter answers. I start with a brief history of sirens, which serve as a guiding analogy. Then, I present current debates in consciousness science. I examine and reject two tempting views of illusionism's implications for consciousness science: the *radical* view and the *quietist* view. Then, I describe illusionism's actual implications. I start with straightforward ones, before arguing that illusionism challenges deep implicit presuppositions of consciousness science, regarding the unity and determinacy of phenomenal consciousness, and the informativity of scientific claims about consciousness. From this, I draw recommendations. I conclude by explaining how the notion of paradigm shift illuminates the opposition between realism and illusionism.

8.1. Consciousness and the Sirens

For centuries, Europeans—laypeople, seafarers, scholars—believed in the existence of *sirens* or *mermaids*: marine creatures with human-like torsos and heads, human-like voices, and a fishy caudal fin. Christopher Columbus reported sighting three mermaids ("*sirenas*") in his first voyage to America, stressing that "they are not so beautiful as they are painted, though to some extent they have the form of a human face." He claimed to have seen them already in Africa, on the coast of Guinea.[1]

[1] See the entry for the 9th of January 1493 (Columbus 2010).

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0009

We now think sirens do not exist. Scientists recognize various species in the order *sirenia*, including the West Indian manatee (*Trichechus manatus*), which Columbus probably saw in America, as well as the African manatee (*Trichechus senegalensis*), which he probably saw in Africa,[2] but sirenians are not human-like—not more than other marine mammals. It is only when seen rising from the water, from a distance, that they visually appear somewhat human-like. The impression seems sufficiently robust to have led to various independent sightings of human-like "sirens" by Europeans seafarers throughout centuries, to have motivated their placing in an order called "*sirenia*" in the 19th century, and it might have generated beliefs about their human-like nature in other times and places.[3]

When did people stop taking sirens seriously? Thomas Bartholin, a prominent Danish physician of the 17th century, who first described the human lymphatic system, seems to have seriously considered the existence of human-like sirens. One century later, Carl Linnaeus himself took reports of sirens' sightings sufficiently seriously to investigate the issue. His research led him to nothing more than a 1-meter-long amphibian he called *Sirenia lacertina* in 1766, which does not bear any resemblance to humans, but is still now taken to belong to a genus called "siren," within the order of *sirenidae* (not to be confused with *sirenia*, which are mammals). By the beginning of the 19th century, no one was taking human-like sirens seriously anymore.[4]

The history of sirens helps illusionists understand what consciousness science should be. Like sirens, phenomenal consciousness does not exist, but seems to exist. Still, *real* and ordinary cognitive and neural processes trigger *real* illusions of phenomenal consciousness, just as *real* and ordinary manatees triggered *real* illusory sightings of human-like sirens. Finally, like the ancient *mythical* sirens, phenomenal consciousness is tempting. Scientists must resist its appeal, the same way legendary seafarers had to resist the sirens' songs.

I will return to the sirens. For now, turn to consciousness science.

[2] Columbus (2010, 154, footnote 1).

[3] Sirens understood as part-human part-fish creatures (i.e., mermaids) appear in many European national folklores. At least in one non-European case, they are associated with manatees: in Nigerian folk religions, *Trichechus senegalensis* is referred to as a Sea-Goddess ("Maami Water") and venerated as such (Orijemie 2021). Greek Homeric sirens were not part-fish, but part-birds.

[4] For more on this, see Scribner (2017), and the tongue-in-cheek Banse (1990).

8.2. Consciousness science

Phenomenal consciousness seems mysterious: obvious, irreducible, and undetectable. Consciousness scientists, unlike philosophers, often sidestep at least one of these mysteries. They avoid the hard problem and set aside the apparent irreducibility of consciousness. They focus on identifying the *cognitive and neural basis* of conscious experience in humans, and remain neutral on whether consciousness reduces to this basis.[5] This has not prevented the emergence of numerous competing views of the basis of consciousness, with little consensus in sight. In a recent non-exhaustive review, neuroscientist Anil Seth and philosopher Tim Bayne count more than twenty-two theories.[6]

Drawing largely on Seth and Bayne's review, we can focus on four of the most influential families of theories. First, *Higher-Order Theories*—HOTs.[7] For HOT, mental states are phenomenally conscious when targeted by a meta-representational state of a certain kind. Second, *Global Workspace Theories*—GWTs.[8] For GWT, mental states are conscious when they enter a global workspace, which makes their content available to a wide range of consumer systems (motor systems, memory, decision-making, reports, etc.). In one influential version, the *Global Neuronal Workspace Theory* (GNWT), this global workspace is identified with specific neural mechanisms, notably involving the pre-frontal cortex (PFC).[9] Third, *Local Recurrence Theory*—LRT.[10] For LRT, perceptual states are conscious when re-entrant processing of a certain kind occurs within sensory areas, without the need for global broadcast or meta-representation. Finally, *Integrated Information Theory*—IIT.[11] IIT claims that a system's degree of consciousness corresponds to the degree (noted Φ) to which states of this system irreducibly integrate information—i.e., carry more information than the sum of the information carried by states of its parts.[12]

[5] Some talk of "neural correlates of consciousness" (NCC): the "minimum neural mechanisms jointly sufficient" for any specific phenomenally conscious experience (Koch et al. 2016, 308; see also Crick and Koch 1990; Chalmers 2000). The *basis* of consciousness (Wu 2018) is an *explanatory* correlate of consciousness, which can fail to be *fully* explanatory. Both "correlate" and "basis" are neutral on irreducibility questions.

[6] Seth and Bayne (2022).

[7] Rosenthal (2005); Lau and Rosenthal (2011); R. Brown et al. (2019); Lau (2022).

[8] Baars (1988).

[9] Dehaene and Naccache (2001); Mashour et al. (2020).

[10] Lamme (2006).

[11] Tononi (2012); Tononi et al. (2016).

[12] Seth and Bayne also give a prominent place to Predictive Processing (PP), but they note that it is less a theory "of" consciousness than a "theory for consciousness": different views use PP to tackle

I did not list illusionist scientific theories of consciousness here. Mainstream realist theories are a more convenient starting point to draw illusionism's implications for consciousness science. If we had to classify scientific proposals made by illusionists, we could say that Dennett's "multiple drafts"[13] and "fame in the brain"[14] views are close cousins of GWT,[15] while Graziano's AST has affinities with HOT. Since, as I will argue, illusionism deeply changes the way consciousness science should be done, any such classification must be taken cum grano salis.

The four major families of theories disagree on the *nature* of the basis of consciousness and on its *physical localization* in the brain. Both HOT and GWT view activations in the PFC as critical to consciousness in humans, while LRT and IIT do not assign such a central role to the PFC and emphasize the role of posterior, early sensory areas.

These theories compete, sometimes brutally. For instance, in October 2023, dozens of prominent consciousness researchers signed an open letter calling one of these four approaches, IIT, "pseudoscience."[16] Few fields feature such conflicts about the scientific character of one of their most influential theories.

Scientists not only disagree about the basis of phenomenal consciousness, but also, relatedly, about the *extension* of phenomenal consciousness in our mental lives and its *distribution* across beings. Regarding the extension question, for instance, researchers persistently disagree on whether all of our phenomenal states are also access-conscious. This *overflow* debate is still ongoing,[17] with no clear resolution in sight. A related debate opposes *sparse* conceptions of consciousness, in which only attended items are phenomenally experienced, and *abundant* conceptions, in which many items are phenomenally experienced outside the focus of attention.[18] Depending on how one views the relations between access and attention, this might come down to the overflow debate.[19]

consciousness. For a skeptical look at PP's contribution to consciousness science, see Schlicht and Dolega (2021).

[13] Dennett (1991).
[14] Dennett (2006).
[15] Dennett (1991; 2006).
[16] Fleming et al. (2023).
[17] Block (2007); Phillips (2011).
[18] Schwitzgebel (2011, chap. 6).
[19] Another unsolved extension debate concerns the existence of cognitive phenomenology, mentioned previously, which mostly interests philosophers.

Regarding the distribution question,[20] disagreement abounds. As Marian Dawkins puts it:

> Proposals for membership of the "consciousness club" include that it is for humans only (Macphail 1987), for language users only (Rosenthal 1993, 2005), for humans and apes only (Bermond 2001), for all mammals (Boly et al. 2013), for mammals and birds (Seth et al. 2005), for mammals, birds and reptiles but not fish or amphibia (Cabanac et al. 2009), for all vertebrates including fish (Denton et al. 2009; Mashour & Alkire 2013; Braithwaite 2010; Sneddon 2019), for all vertebrates and a few invertebrates such as octopuses (Tye 2017), for many invertebrates especially insects and crustacea (Klein & Barron 2016; [Bronfman et al., 2016]) or that it should be for all living things, including plants (Margulis 2001). It has even been claimed that consciousness is everywhere, even in inanimate objects ([Chalmers 2015]; Kastrup 2018).[21]

There are also persisting debates about human infants or fetuses,[22] cerebral organoids—roughly, lab-grown models of brain parts—,[23] isolated brain hemispheres,[24] plants,[25] and, of course, artificial systems.[26]

On all these questions—the nature and localization of the basis of phenomenal experience, and the extension and distribution of phenomenal consciousness—there is radical and persistent disagreement. This occurs even if irreducibility concerns are set aside. Why?

A natural explanation is that consciousness scientists face the *detection* (or *measurement*) problem.[27] The only detection methods that seem antecedently adequate for phenomenal consciousness are first-personal: they require subjects to introspect and report (e.g., verbally) their introspection.[28]

[20] Allen and Bekoff (1997).
[21] Dawkins (2021, 6).
[22] Bayne et al. (2023).
[23] Zilio and Lavazza (2023).
[24] Gauvry and Theodor (2024).
[25] Segundo-Ortin and Calvo (2022).
[26] Butlin et al. (2023). There are also philosophical debates regarding the presence of consciousness in group entities (Schwitzgebel 2015, 2016a; Kammerer 2015).
[27] Browning and Veit (2020); Seth and Bayne (2022).
[28] Consciousness researchers, asked about the "best ways to check whether stimuli are consciously perceived," first answered "subjective report." The two other most popular answers are closely related (Francken et al. 2022).

Other methods (e.g., neural measures) are hard to justify antecedently.[29] However, subjective reports generate issues.

First, they converge in simple cases: all sighted subjects report visual experiences of brightly colored, stable stimuli in front of them. However, they do not converge in difficult cases relevant to settle scientific debates. Subjects make contradictory or uncertain reports on cases relevant to the overflow question or the sparse vs. abundant question.[30] It was also argued that *all* possible operationalizations of the probing of subjective reports exhibit ineliminable biases and yield contradictory measurements reflecting these biases: no non-arbitrary choice can be made between these operationalizations.[31] Clever bias-free subjective methods have been proposed,[32] but whether they genuinely detect consciousness is unclear.[33]

Second, setting this aside, subjective reports remain unavailable for non-humans and cannot directly help with the distribution question. Animals cannot produce verbal reports, nor can they produce non-verbal reports unambiguously bearing on phenomenal experiences rather than external stimuli.[34] The same applies to infants, fetuses, plants, or cerebral organoids. Some AIs [e.g., Large Language Models (LLMs)] produce statements *resembling* subjective reports, but even on the assumption that their statements are meaningful,[35] they are unreliable, at least for current models,[36] and there are principled reasons to think that they might remain so.[37]

So, only subjective methods seem antecedently adequate to study phenomenal consciousness, but these methods are seriously limited. This is the detection problem, which I think is linked to the mysteries of consciousness

[29] This does not mean that detection methods that do not rely on subjective reports—"no-report" methods (Tsuchiya et al. 2015)—cannot be used. These methods infer consciousness from third-personally investigable features taken to be associated with consciousness. However, this association must itself be justified, and subjective reports seem unavoidable to do so. For more on no-report methods, see Block (2019) and Schlicht (2018).

[30] Schwitzgebel (2011) makes this case at length. Note that most subjects (but not all of them), when probed, describe their visual experiences in ways contradicting the overflow hypothesis (Cova et al. 2021).

[31] Irvine (2012).

[32] Galvin et al. (2003); Maniscalco and Lau (2012).

[33] Irvine (2013); Rosenthal (2019b). For an optimistic take, see Michel (2023).

[34] See Browning and Veit (2023) for an optimistic take, and Kammerer and Frankish (2023a, 257) for a response.

[35] For a skeptical take, see Bender et al. (2021).

[36] For a reasonably optimistic take, see Long (2023); Perez and Long (2023).

[37] See what Jonathan Birch (2024) calls the "gaming problem" regarding LLMs and consciousness-related statements.

I discussed. We grasp phenomenal consciousness as introspectively obvious, but also directly undetectable from the third-person perspective. Consequently, we give an ineliminable but problematic role to subjective reports.

This detection problem might explain why consciousness science is conflictual, with so little consensus on key issues. Some might hope that this lack of consensus only reflects the relative youth of the field: on a popular narrative, consciousness science really took off in the 1990s. Close attention to history suggests that this is wrong. As Matthias Michel convincingly argued,[38] the scientific study of consciousness began as early as the mid-19th century. Researchers back then had persistent debates, structurally similar to ours, about the localization of the basis of conscious experience and the extension question.

Many consciousness scientists hope that the right approach will ultimately allow us to build consensus in the field. One strategy that received attention recently is the iterative approach, also known as the natural-kind approach.[39] The idea is that scientists should start with imperfect indicators of consciousness (including subjective reports) and use them to hypothesize about the phenomenon, ideally latching on, albeit imperfectly, to something like an underlying natural kind. This should allow them to refine their indicators, run new experiments using these indicators, which should, in turn, help them formulate better hypotheses, etc. The hope is that, through this iterative process, they will zero in on the natural kind corresponding to phenomenal consciousness and reach a satisfying scientific theory of consciousness as well as robust detection methods.

Many consciousness scientists remain hopeful, for instance, because they trust this natural kind approach. Still, contemplating the current state of the field from afar, one appreciates the relative absence of consensus as well as the radical and entrenched nature of theoretical oppositions. This makes skepticism tempting: What if consciousness lies, forever, beyond the reach of science?[40]

[38] Michel (2019).

[39] It might be the most popular approach for optimistists about consciousness science (Bayne et al. 2024; Bayne and Shea 2020; Birch 2022b; Mckilliam 2025; Shea 2012).

[40] Full-blown skepticism is rare. See Schwitzgebel (2011) for qualified skepticism on the extension question. See Block (2002) and Prinz (2003) for qualified skepticism on the distribution question.

8.3. Two tempting views of illusionism's implications

What does illusionism imply for consciousness science? Here is a first tempting idea: illusionism implies that consciousness science is moot. Since phenomenal consciousness does not exist, it lacks a neural or cognitive basis. The extension and distribution questions have simple answers: no mental states or beings are phenomenally conscious. This is the *radical* view: illusionism provides a radical negative answer to all the main questions of consciousness science, rendering it pointless, comparable to cryptozoology and its study of human-like sirens, dragons, or unicorns.

Some will welcome the radical view. Consciousness science is riddled with problems: if illusionism justifies that we get rid of it, so much the better! Others will remain suspicious and think that we should be wary of philosophical view implying that an entire scientific field is ready for the scrapheap.

Against the radical interpretation, illusionism does not deny functionally conceived forms of consciousness, like access-consciousness or self-consciousness. Consciousness scientists can study those without being siren-chasers. Still, as the study of phenomenal consciousness is a primary aim of current consciousness science, proponents of the radical view can maintain that illusionism implies that core parts of consciousness science are moot.

Some could retort that current consciousness science does *not* really aim at studying phenomenal consciousness. Prominent researchers, often proponents of GWT, refuse to distinguish between access-consciousness and phenomenal consciousness, and claim to study "conscious access."[41] However, one can accept the *conceptual* distinction between access and phenomenality while denying the real and empirical distinction. Many presentations of GWT in fact seem to presuppose a conceptual distinction, while asserting the empirical identification. Stanislas Dehaene and Lionel Naccache, two prominent neuroscientists who developed GNWT, claimed that:

> global availability of information through the workspace *is* what we subjectively experience as a conscious state.[42]

[41] Mashour et al. (2020).
[42] Dehaene and Naccache (2001, my emphasis).

Peter Carruthers, who accepted GWT after years of defending HOT, stated that:

> phenomenally conscious states *are* those that are access-conscious.[43]

So, even GWT plausibly studies phenomenal experience and *identifies* it with conscious access. Moreover, as a matter of sociological fact, most consciousness scientists accept, not only a conceptual distinction, but also a real distinction between phenomenal and access-consciousness.[44] All this makes it hard to deny that studying phenomenal consciousness is a core aim of consciousness science. The radical view remains tempting: illusionism makes core parts of consciousness science moot.

This radical view faces a problem. Illusionists deny phenomenal consciousness, but recognize tightly related phenomena as real. *Illusions* of phenomenal experiences are real. These illusions are triggered by real states, which they track and mischaracterize as phenomenal: *quasi-phenomenal states*. So, two real phenomena—the illusion of phenomenality and quasi-phenomenality—are closely related to phenomenal consciousness but distinct from it. These phenomena are also conceptually distinct from access-consciousness or self-consciousness. About them, the standard questions of consciousness science arise. We can ask about the nature and localization of their *basis*. We can ask about their *extension* in our mental lives and their *distribution* across beings.

These questions resemble those usually asked about phenomenal consciousness. For instance, take the quasi-phenomenal extension question. Which of our mental states are quasi-phenomenal? That is: Which are of the same type as states tracked and mischaracterized as phenomenal, notably by phenomenal introspection? Answering this question, we can reformulate an overflow debate and wonder whether all quasi-phenomenal states are also access-conscious. This question, prima facie, is open.

For some proponents of GWT, we should only study consciousness understood as *conscious access*, since using a notion of consciousness distinct from access makes it "immune to scientific investigation,"[45] or "leads down a slippery slope of dualism."[46] Illusionism does *not* have this implication.

[43] Carruthers (2019, 103).
[44] Most consciousness researchers think phenomenal consciousness overflows access (Francken et al. 2022, 4).
[45] Kouider et al. (2012).
[46] Dehaene (2014, 10).

Given illusionism, our concept of quasi-phenomenal consciousness refers to something real and potentially investigable, while being distinct qua concept from the concept of access-consciousness. (Of course, it is not entirely divorced from the concept of access broadly understood, and certainly not from functional concepts.)[47] The two concepts might refer to empirically distinct phenomena: quasi-phenomenality could overflow access.

Other example: take distribution questions. We can ask a *quasi-phenomenal distribution question*: Beyond humans, which beings have mental states of the same type as those that, in us, are introspectively tracked and mischaracterized as phenomenal? We can also ask the *illusion distribution question*: Which beings have the illusion of phenomenal consciousness? Again, given illusionism, these questions are meaningful and seem open.

So, while illusionism gives radical negative answers to questions regarding *phenomenal consciousness*, it leaves other resembling questions open. These open questions pertain to real, closely related, but distinct phenomena: the illusion of phenomenality and quasi-phenomenality. This suggests that illusionism does not make consciousness science moot or pointless: the radical view is wrong.

This can easily lead us to another conception: the *quietist* view. For the quietist, illusionism has no significant implications for the science of consciousness. Illusionists *can* ask standard consciousness science questions, as long as they ask them about quasi-phenomenal consciousness or the illusion of consciousness. All that illusionism requires of consciousness scientists is verbal adjustments specifying that they discuss these phenomena, not phenomenal consciousness.

An attractive approach, for a quietist, is to take *one* of these two phenomena as the privileged heir of phenomenal consciousness in an illusionist framework. The best candidate is probably quasi-phenomenality, about which standard consciousness questions can be asked in a way prima facie similar to how they are ordinarily asked about phenomenal consciousness. Consciousness scientists can then claim that they do not study phenomenal consciousness, but the form of consciousness that wrongly appears phenomenal, and insist that this does not otherwise alter their work and methods. Illusionism would be a purely philosophical position, devoid of

[47] It satisfies the idea that scientists should not divorce consciousness from function (M.A. Cohen and Dennett 2011).

scientific relevance, which leaves consciousness science intact. Similarly, metaphysicians debate whether arrangements of atoms constitute genuine objects such as plants or tables, but botanists and carpenters safely ignore their conclusions.

The quietist and radical views are simple versions of, respectively, the substitutionist and abolitionist stances toward the "Now what?" question in the field of consciousness science. Both views, however, are misguided. Illusionism does not make consciousness science moot. It does not leave it untouched either. It has serious, non-simplistic, and non-trivial implications for consciousness science, to which I now turn.

8.4. Straightforward implications of illusionism

Illusionism has four straightforward implications. First, by denying phenomenal consciousness, it denies the entity about which the hard problem and the explanatory gap arise. This is not trivial: a clear majority of consciousness researchers[48] and philosophers[49] think that there is a hard problem or an explanatory gap. This denial is impactful: the fact that a scientific theory promises to address the hard problem or the explanatory gap more effectively than others is no longer an advantage, contrary to what is often believed. For instance, Ned Block claimed that the biological approach to consciousness (to which versions of LRT might belong) "handles the explanatory gap better than [. . .] the global workspace or higher order views." He saw this as an advantage of the view, by contrast with other views, like HOT, which are only attractive to "people who do not agree that there is an explanatory gap."[50] Defenders of IIT also claimed that their view "addresses the hard problem in a new way."[51] Given illusionism, these putative features of LRT or IIT do not count as advantages.

Second, illusionism has anti-skeptical implications. Examining current conflicts in consciousness science, one might fear that its object lies beyond the reach of science. Illusionism denies that phenomenal consciousness, with all its mysteries, exists. This should give us higher antecedent confidence that

[48] Francken et al. (2022).
[49] Chalmers and Bourget (2020).
[50] Block (2009, 1111–14).
[51] Tononi et al. (2016, 450).

all real consciousness-related phenomena are within the bailiwick of science, even if they do not reduce, conceptually or empirically, to conscious access. Of course, this does not tell us *how* scientific questions about these phenomena can be answered.

Third, for illusionism, phenomenal introspection is mischaracterizing: it misrepresents the quasi-phenomenal as phenomenal. This reduces the attractiveness of approaches using introspection to establish basic claims about the nature of consciousness. For instance, IIT relies on "phenomenal axioms"—basic statements about consciousness—supposedly established by "drawing on introspection and reason."[52] Illusionism suggests that we should *not* start with such introspectively justified axioms.[53]

Fourth, illusionism suggests a change of focus. We should stop studying phenomenal consciousness, which does not exist, but pay attention to our *representations* of phenomenal consciousness, which shed light on the two real phenomena most related to phenomenal consciousness: quasi-phenomenality and the illusion of phenomenality. Research on the illusion problem and the meta-problem of consciousness should become central to consciousness research. This is hardly the case currently. Mainstream consciousness science treats the meta-problem, at best, as an afterthought, when it does not ignore it.[54] Solving the meta-problem is rarely seen as a core goal of consciousness science, except maybe in the illusionist or illusionist-friendly camp.[55]

Illusionism thus also suggests giving more attention to phenomenal introspection, not just as the basis of a *method* to study consciousness, but as a representational process to be studied in its own right. Introspection as a method has received much attention in cognitive science,[56] and even if this method raises doubts,[57] or at least calls for caution.[58] Introspection as a process to be studied for itself has been relatively neglected in cognitive

[52] Albantakis et al. (2023)

[53] McQueen (2019) defends illusionist IIT, about quasi-phenomenal consciousness. I doubt that the adapted "axioms" of his illusionist IIT can make postulates similar to those of standard IIT attractive.

[54] Chalmers (2018, 35–40) convincingly argued the meta-problem poses a challenge for theories of consciousness, notably for IIT and, to some extent, for GWT and HOT, as these theories are not well-equipped to solve it in a manner that coheres with their claims about consciousness.

[55] Humphrey (2011); Graziano (2013); A. Clark et al. (2019).

[56] Both historically (Titchener 1901; Wundt 1874) and more recently (Ericsson 2003; Hatfield 2005; Hurlburt and Heavey 2001; Petitmengin 2006), after a long period of disappearance (W. Lyons 1986).

[57] Schwitzgebel (2011); Irvine (2013; 2021).

[58] Feest (2014); Timmermanns and Cleeremans (2015); Spener (2024).

science, at least when it comes to *phenomenal* introspection.[59] Empirical research on introspection or metacognition in humans[60] and animals[61] focused mostly on representations of knowledge, confidence, decisions, etc.

8.5. Consciousness science's presuppositions

Illusionism has deeper implications for consciousness science, aside from these straightforward ones. It challenges consciousness science's presuppositions.

Standard consciousness science looks for the basis, extension, and distribution of phenomenal consciousness. Doing so, it relies on presuppositions. I want to zero in on three of them. I will not argue for this, but I think that they are widely (though not universally) shared, even if they usually remain implicit.

First, standard consciousness science often implicitly treats phenomenal consciousness as a reasonably unified phenomenon, corresponding to a single and unified set of mental states. Second, standard consciousness science often implicitly understands claims about the basis, extension, and distribution of phenomenal consciousness as having a *determinate* truth-value: as determinately true or false. For instance, scientists often assume that a given creature either has phenomenal experiences or does not. Third, standard consciousness science often implicitly assumes that knowing about the basis, extension, and distribution of phenomenal consciousness would impart us with substantive and important knowledge about our minds and other minds. It assumes that claims about phenomenal consciousness are highly informative. For instance, scientists assume that we would learn something substantive and important about a shrimp or a robot when learning that they are phenomenally conscious.

These presuppositions concern, respectively, the *unity* of the object of consciousness science, the *determinacy* of the truth-value of the claims made about it, and the *informativity* of these claims. Illusionism challenges all three of them. I will show how this impacts consciousness science.

[59] With some exceptions (e.g. Overgaard et al. 2006).
[60] Nisbett and Wilson (1977); Flavell (1979); Lai (2011); Fleming (2021).
[61] Beran (2019); Carruthers and Williams (2019).

8.6. Consciousness sawed in half: Against unity

Standard consciousness science implicitly assumes that phenomenal consciousness is a unified phenomenon. This phenomenon might comprise multiple distinct properties coming in degrees, making it multi-dimensional,[62] but there remains an assumption that it is sufficiently unified: a number of true and interesting scientific claims can be made about *phenomenal consciousness* as such.

Illusionism answers questions about phenomenal consciousness in a radical and negative manner, but leaves related questions open. However, these related questions bear on *two* phenomena, two candidate heirs of phenomenal consciousness given illusionism: *quasi-phenomenal consciousness* and the *illusion of phenomenal consciousness*.

Illusionist consciousness scientists must carefully reflect to decide which of these phenomena they target. Earlier, I claimed that quasi-phenomenality is often the most natural choice, but a choice still has to be made. This has various consequences. Here are two of them.

First, notice that among the four leading theories of the basis of phenomenal consciousness, only *one*—HOT—requires phenomenally conscious states to be represented. GWT, IIT, and LRT do not. HOT is a higher-order view of consciousness; the others plausibly count as 'first-order' views.

In the philosophical literature, the first-order/higher-order distinction often applies to representationalism: first-order representationalism identifies phenomenal consciousness with certain first-order mental representations of external or bodily features, while higher-order representationalism is similar to HOT. Using this philosophical distinction to map scientific theories is tricky, but the "first-order" label works nicely, at least for LRT and GWT, which view phenomenal experiences as representations that *can* be merely first-order. Standard versions of first-order representationalism, like Michael Tye's PANIC theory,[63] are fairly similar to GWT.

Illusionism implies that consciousness science, instead of phenomenal consciousness, faces these two phenomena: quasi-phenomenal consciousness and the illusion of phenomenal consciousness. It thus offers a conciliatory perspective on the opposition between first-order and higher-order theories. The *illusion* of phenomenality certainly involves higher-order

[62] Bayne et al. (2016); Birch et al. (2020); Dung and Newen (2023).
[63] Tye (1995).

representational processes: having the illusion requires misrepresenting some mental states (as phenomenal). So, the basis of the illusion must involve higher-order processes, like those described by HOT. On the other hand, quasi-phenomenal consciousness—the type of mental states tracked and mischaracterized by phenomenal introspection—is more naturally, though not necessarily, understood as a process that does not require higher-order representation. First-order views plausibly describe its basis.

Therefore, theories that *conflict* about phenomenal consciousness can become compatible if interpreted as about these two distinct phenomena. This interpretation is naturally suggested by illusionism, which offers a potential conciliatory dissolution of entrenched oppositions. Of course, illusionism does not *require* this dissolution. For instance, properties of first-order processes could explain much about the illusion of phenomenal consciousness—e.g., why some processes are characterized as phenomenal and not others. If so, the correct theory of the illusion of phenomenality includes elements of a first-order view.[64] Conversely, quasi-phenomenal consciousness could turn out to involve higher-order processes.

A quietist can insist that illusionism is not *required* for the dissolution. Prominent non-illusionist proponents of GWT and HOT advocated for this dissolution, claiming that "the word 'consciousness' conflates two types of information-processing computations,"[65] some first-order (conscious access) and some higher-order (the monitoring of conscious access). However, they acknowledged that this raises an objection, later pressed by opponents,[66] as the resulting view seems to "[leave] aside the experiential component." Illusionism precisely gives reasons to leave this component aside as non-existent, thus providing solid ground for this dissolving and conciliatory move.[67]

Here is a second comparable implication of illusionism's challenge to the unity presupposition. The consciousness distribution question received radically conflicting answers, ranging from restrictive ones, which hold that only humans and apes are phenomenally conscious,[68] to liberal ones, on which many beings are conscious—say, up to insects.[69] For illusionism,

[64] Thanks to Matthias Michel for this point.
[65] Dehaene et al. (2017).
[66] Carter et al. (2018).
[67] Graziano et al. (2020) make a similar point, stating that AST "can be understood as a specific unification of [GWT] and HOT"—AST's illusionism playing a key role in this unification.
[68] Carruthers (1992); Bermond (2001).
[69] Klein and Barron (2016).

no one is *phenomenally* conscious, but the illusion distribution question and the quasi-phenomenal distribution question remain open. These two questions plausibly call for different answers. Unsophisticated beings could have states like those tracked by phenomenal introspection in us, and be quasi-phenomenally conscious, but lack the illusion of phenomenality. Sophisticated beings very different from us could have an illusion of phenomenality without being quasi-phenomenally conscious, if the states mischaracterized as phenomenal in them are significantly different.

This also suggests a conciliatory dissolution of rival views. Among animals, restrictive views seem attractive for the *illusion* distribution question—say, views reserving the illusion to humans. Liberal views are more attractive for the quasi-phenomenal distribution question. Sophisticated AIs with the right metacognitive capacities might join the illusion club, but maybe not the quasi-phenomenal club. Again, illusionism does not mandate this conciliation, but it makes it natural. It is not required for it, but it grounds it theoretically.

It helps to think of sirens. Facing the question "Where are sirens?", we must draw distinctions. First, 'genuine' sirens (human-like mammal creatures with caudal fins) do not exist anywhere. Second, creatures triggering illusory sightings of sirens, such as manatees, really exist.[70] Manatees, for instance, appeared millions of years ago and now live in coastal areas in America and Africa. Third, *illusory sightings* of sirens occurred abundantly until the 17th century in various places,[71] and became rarer in the 18th and 19th centuries before disappearing. Illusionists need comparable distinctions to answer the distribution question, which might lead to conciliatory dissolutions of entrenched oppositions.

8.7. Consciousness and indeterminacy

A. Indeterminacy and the distribution of quasi-phenomenality

Illusionism not only changes consciousness science by replacing one object and one set of questions with two. It also affects answers to these questions by challenging the determinacy presupposition.

[70] Do creatures other than manatees, like sirenidae (e.g., Linnaeus's *Sirenia lacertina*) also count? The question might fail to have a determinate answer—see the next section.

[71] For instance, see Scribner (2017, 511).

Consider the quasi-phenomenal distribution question. Suppose we correctly describe the mental states tracked by phenomenal introspection in us—our quasi-phenomenal states. For instance, suppose GNWT correctly describes them. Our quasi-phenomenal states are GNWT states.

Our quasi-phenomenal states, therefore, belong to various types. First, they belong to differently determined *functional* types. They are globally broadcast representations (a), globally broadcast representations outputted by various modules working in parallel (b), globally broadcast representations available for memory subsystems (c), and for verbal reports (d)—etc. Second, they belong to differently determined *neural* types. They are chemical and electrical activities of biological neurons (α), and neural activities involving activations in the frontal and parietal cortices (β), as well as neural activities involving activations in some more specific networks in the PFC (γ), etc. Being GNWT states, they belong to a combined specific type: the type of states with a, b, c, d, etc., as well as α, β, γ, etc. Finally, they belong to countless other types not mentioned by the theory—e.g., they are all things that exist in the Solar System! This is unsurprising. Everything belongs to various types.

Many beings may thus have states with some, but not all, of the properties of our quasi-phenomenal states. Some future AIs might have globally broadcast representations outputted by a variety of modules working in parallel and available to the suitable subsystems (with a, b, c, and d) but lack biological neurons (without α, β, γ, etc.). Various animals arguably have globally broadcast representations available to many subsystems (with a, b, c), but inaccessible to report (without d). Fish, lacking a cortex, might have neuron-based globally broadcast representations (with a and α) without cortical activity (β and γ)—and so on.

With this in mind, consider again the quasi-phenomenal distribution question: Which beings have quasi-phenomenal states? That is: Which beings have states of the same type as those tracked by phenomenal introspection in us? Plausibly, the best answer to this question lacks a determinate truth-value. Many other beings' mental states are, in many respects, but not others, of the same type as those tracked by phenomenal introspection in us. Whether they are quasi-phenomenal is indeterminate.[72]

[72] This is adapted from arguments for the indeterminacy of phenomenal consciousness developed by weak *realists* (Carruthers 2019; Papineau 2002). These arguments are made from a type-B materialist perspective, on which we do not refer to phenomenal consciousness via an informative conception. The argument can naturally bear on quasi-phenomenal consciousness in an illusionist

Again, it helps to think of sirens. Suppose an alternative reality in which 19th-century zoologists discover that there are no human-like sirens. Zoologists call beings tracked by our illusory sightings of sirens "quasi-sirens." They realize that these sightings typically track marine mammals found in the Caribbean and Central American coasts (*Trichechus manatus*). Traveling the world, they discover other similar creatures. Some, in Africa, resemble *Trichechus manatus*, and locally trigger illusory sightings of human-like creatures (*Trichechus senegalensis*). Now imagine that, in this alternative reality, zoologists discover species that do not exist in ours: e.g., *Trichechus impar*, which are manatees functionally and evolutionarily similar to *Trichechus manatus*, with superficial differences in the rostral areas making it so that they never appear human-like and cannot trigger illusory sightings of sirens. They also find species of sharks, very different from *Trichechus manatus*, with a human-like look, causing sightings of human-like creatures. Finally, mischievous 21st-century geneticists engineer marine crocodiles to resemble humans and release them in isolated areas where beliefs in sirens are widespread, triggering sightings of sirens.

In this scenario, plausibly, many creatures have an indeterminate quasi-siren status. They are *like* prototypical quasi-sirens (*Trichechus manatus*) in some significant respects and unlike them in others. This situation seems analogous to what happens with quasi-phenomenal consciousness across beings: its distribution, plausibly, is also indeterminate.

The indeterminacy at play here is radical. The issue is not simply that the extension of quasi-phenomenal consciousness is fuzzy, and that beings close to the fuzzy border fail to be determinately quasi-conscious or not.[73] The issue goes beyond borderline cases: it is not a situation with a fuzzy border, but one with *many distinct valid candidate borders*—all potentially fuzzy, but we can set this aside here. Each candidate border carves the territory differently, without a fact of the matter as to which is the right true border. With only one fuzzy border, we could (fuzzily) delineate the border area and

framework, since quasi-phenomenal consciousness is also grasped without a highly informative conception—as the type of states tracked and mischaracterized by phenomenal introspection. Tightly connected issues include the "specificity problem" (Shevlin 2021) and the "demandingness dilemma" (Birch 2022b) arising about the distribution question. Realists who developed this indeterminacy argument endorse very weak forms of realism, close to illusionism in spirit (C. D. Brown and Papineau 2024; Carruthers 2019).

[73] Something that can be argued with a sorite-like argument (examined in Cutter 2017; endorsed in Schwitzgebel 2023)

contain the issue, which would concern a limited class of beings. With many distinct valid candidate borders, a lot more beings are concerned. The issue cannot be easily contained.[74]

This sort of indeterminacy is avoided if, among the types our quasi-phenomenal states belong to, one stands out. It could be a *uniquely* salient and important specific natural kind corresponding to a more important joint in nature, better supporting inductive generalizations, serving as a reference magnet for our concept.[75] The idea that consciousness corresponds to such a natural kind is key to the natural kind approach mentioned previously. The type standing out could also be a single specific type our phenomenal introspection has the *function* to track, for instance, if phenomenal introspection was selected because it tracked it. If Meta-Cartesianism is correct, it could be the property of being a reliable non-inferential belief-formation process. In such cases, quasi-phenomenal consciousness corresponds to a single specific type, whether we know it or not. Beings with states of this type are determinately quasi-phenomenally conscious; others are not.

Things *could* turn out this way. Is it plausible? First, note that, if GNWT is correct about our quasi-phenomenal states (and the same applies to other theories, like HOT), our GNWT states belong to functional types (a, b, c, etc.) and neural types (α, β, γ, etc.). However, as Jonathan Birch convincingly argued, since these neural properties are supposed to *realize* these functional properties, we can account equally well for the success of inductive generalizations made with a concept of quasi-phenomenal consciousness by supposing that it refers to either of the two types.[76] Inductive success does not help determine the reference of the concept. Similarly, the property of being a reliable non-inferential belief-formation process is neurally realized: if phenomenal introspection was selected because it tracks these processes, then it was also selected because it tracks their neural realizations.

[74] On "mild" versus "radical" indeterminacy, see Birch (2022a, 802–6).

[75] Shea (2012); Simon (2020). For more on reference magnetism, see Lewis (1984); Sider (2011). This argument for the indeterminacy of consciousness resembles the determinacy challenges faced by teleological theories of naturalistic content (Artiga 2021; Fodor 1990; Neander 2017), sometimes addressed by appealing to natural kinds (Martínez 2013).

[76] Birch (2022a, 806–9). For Birch, to establish that successful inductive generalizations made with our concept is supported by its reference to the functional type rather than the neural type (the other way around is impossible since the right neural types realize the functional types), we must suppose that our concept supports successful inductions when a being instantiates the functional type without the neural type. However, presupposing that these inductions are successful in these cases *precisely* requires presupposing that our concept refers to the functional type—which is the point in question.

Second, while many researchers confidently bet that consciousness (usually meaning *phenomenal* consciousness) corresponds to a single specific natural kind, illusionism gives a debunking explanation of this confidence. Phenomenal consciousness, introspectively, appears special and distinctive. It appears uniquely mysterious. As a result, it seems natural to infer that it must form a single specific kind, cutting nature at its joints, and whose presence is determinate. However, for illusionists, phenomenal consciousness does not exist. They have no particular reason to think that what really exists, and is mischaracterized as phenomenal, is special and distinctive. Illusionism debunks our intuition that it is special and distinctive, and therefore, our intuition that it forms a single specific natural kind, as stemming from implicit realist presuppositions. So, given illusionism, it is relatively less likely that quasi-phenomenal consciousness corresponds to a single natural kind. This also impacts the prospects of illusionist versions of the natural kind approach.

Third, think of cases other than consciousness. Many psychological processes superficially *appear* to constitute a single unified kind, but further research suggests that they might rather correspond to a variety of loosely connected, distinct kinds, making the distribution of the unspecified process indeterminate. Take *memory*. Superficially, it looks like a single psychological faculty.[77] However, contemporary cognitive science suggests that, even limiting ourselves to humans, it corresponds to various distinct kinds: short-term and long-term memory, episodic and semantic memory, explicit and implicit memory, etc.[78] Immunologists also speak of *immune memory* (grounded in so-called memory cells); various single-cell organisms also seem able to "remember" sequences of events[79]; slime molds (brainless, single-celled organisms) might enjoy a form of externalized memory.[80] Humans themselves might have "extended" memories and remember things written in their notebooks.[81] As a result, it is plausible that, for many Xs, unspecified claims like "X is capable of memory" lack a determinate truth-value.[82]

[77] Locke apparently treats it as a single unique faculty in the Essay, book II, chapter X (Locke [1689] 1975).
[78] Baddeley (1999).
[79] Saigusa et al. (2008).
[80] Sims and Kiverstein (2022).
[81] A. Clark and Chalmers (1998).
[82] For arguments that memory *tout court* is not a natural kind, see Michaelian (2011); Rupert (2013). See Boyle (2022) for a defense of kind pluralism on the topic of episodic memory, in which the delineation of the relevant natural kind(s) is irremediably dependent on the theoretical and practical interests of various groups of researchers. See Colaço (2022) for a tentative methodological defense of the treatment of memory tout court as a broadly defined natural kind.

The case of memory does not seem exceptional, making it plausible that similar outcomes could occur for quasi-phenomenal consciousness.

All this supports the claim that the distribution of quasi-phenomenal consciousness across beings is not just unknown, but probably indeterminate.

B. Indeterminacy, and the basis and extension of quasi-phenomenality

It seems less probable, but the extension of quasi-phenomenal consciousness in our mental lives could also be indeterminate. To see why, drop the assumption that our quasi-phenomenal states are GNWT states. Still, suppose that we know everything there is to know about how and when phenomenal introspection is triggered.

Plausibly, some types of mental states trigger phenomenal introspection unsystematically. Consider high-level cognitive processes, like understanding a sentence in our native language. If my hypotheses in Chapter 6 are correct, it *sometimes* triggers phenomenal introspection: more systematically so than the thought that 13 is prime, but less so than a perception of red. A plausible conclusion is that perceptions of red are determinately tracked by phenomenal introspection, thoughts that 13 is prime are not, while things remain indeterminate for understanding a sentence. The upshot is that this last process has an indeterminate quasi-phenomenal status. This impacts the cognitive phenomenology debate, reinterpreted as concerning quasi-phenomenology: some cognitive states would not be determinately quasi-phenomenal or not.

Different properties could also play a role in the causal chain leading to phenomenal introspection. Suppose re-entrant processes within sensory areas (LRT states) tend to cause GNWT states, which then tend to cause phenomenal introspection. If so, both LRT and GNWT states tend to trigger phenomenal introspection.[83] While this is not straightforward,[84] it could be that there is no fact of the matter, ultimately, as to whether phenomenal introspection tracks LRT states, GNWT states, their conjunction, or a wider type of states that includes them, etc. The extension of quasi-phenomenal

[83] This is plausible even if one denies that causation is always transitive, a view defended by Lewis (1973a), which faces plausible counterexamples.

[84] This is reminiscent of so-called horizontal determinacy challenges for teleological views of content (Artiga 2021; Neander 2017).

consciousness would be indeterminate. This impacts the overflow debate: assuming also that, as is plausible, LRT states, but not GNWT states, overflow access, the view that quasi-phenomenality overflows access would fail to be determinately true or false.[85]

Finally, the indeterminacy of the extension of quasi-phenomenality arguably entails the indeterminacy of its basis. For instance, if processes of sentence comprehension have an indeterminate quasi-phenomenal status, it is indeterminate whether the neural and cognitive basis of these processes belongs to the basis of quasi-phenomenal consciousness. Similarly, if LRT states have an indeterminate quasi-phenomenal status, the basis of quasi-phenomenal consciousness is indeterminate.

C. Illusionism and indeterminacy

A quietist could point out that this argument for the indeterminacy of quasi-phenomenal consciousness does not require illusionism. Realists have made similar cases for indeterminacy about phenomenal consciousness.

Still, illusionism strengthens the case for indeterminacy. First, while we might have the intuition that *phenomenal* consciousness forms a distinctive and determinate natural kind, illusionism, as I noted, debunks this intuition if reinterpreted as about *quasi-phenomenal* consciousness. Second, an influential objection to the indeterminacy of phenomenal consciousness becomes toothless about quasi-phenomenal consciousness. It is an a priori objection appealing to the *positive inconceivability* of indeterminate consciousness—to the fact that we seem unable to positively represent to ourselves cases of indeterminate consciousness.[86]

Here is a way to present the objection. Plausibly, when a concept C admits of indeterminate cases, we can describe some cases without using C, and intuit that they do not determinately satisfy nor fail to satisfy C. For instance, BALD admits of indeterminate cases. We can think of various cases without using BALD (e.g., consider people with different amounts of hair) and intuit that some are cases of indeterminate baldness.

[85] Schwitzgebel (2011, 179; 2023, 21–22) makes a similar suggestion.

[86] Versions of this objection were developed (Antony 2006; Simon 2017; Tye 2021), usually with a focus on *vagueness* rather than indeterminacy per se.

However, no description yields the intuition that we face a case of indeterminate consciousness. Trying to find such a case, we might think, say, of the confused, dim sensations we experience in the morning as we wake up, but on reflection, we intuit that these experiences are *determinately* phenomenally conscious: they are fleeting, but intuitively, it remains determinate that they are phenomenal experiences. The same applies to other candidate cases of indeterminate consciousness. So, we do not intuitively grasp what indeterminate consciousness is, or could be. Some conclude that it cannot exist.

This objection might have weight for realists. Illusionists can ignore it. Suppose we fail to positively conceive of indeterminate *phenomenal consciousness*. This might show something about phenomenal concepts: e.g., that they do not admit of indeterminate cases. This is no problem for illusionists, who think everything real determinately *fails* to satisfy phenomenal concepts. For illusionists, the key question is whether the concept of *quasi-phenomenal consciousness* admits of indeterminate cases. This objection is silent on the matter. Besides, it seems easy to positively conceive of indeterminate *quasi*-phenomenality, as my earlier descriptions suggest. Suppose phenomenal introspection, in us, tracks and mischaracterizes states with distinct separable properties F, G, H, and I, all of which are comparably salient and significant natural kinds. Imagine a creature possessing states with F and G, but lacking H and I. Arguably, by doing so, you just positively conceived of a creature that is not determinately quasi-phenomenally conscious or not. Intuitively, this creature's mental states do not determinately satisfy (nor fail to satisfy) the concept of *a state of the same kind as those states tracked and mischaracterized by phenomenal introspection in us*. We might struggle to conceive positively of indeterminate phenomenal consciousness, but we do it easily with quasi-phenomenal consciousness. The inconceivability objection against indeterminate phenomenal consciousness does not threaten indeterminate quasi-phenomenal consciousness.[87]

Let us summarize. The distribution of quasi-phenomenality is probably indeterminate. Possibly, its extension and basis are indeterminate too. Illusionism makes this indeterminacy of quasi-phenomenality more likely than the indeterminacy of phenomenality given realism.

[87] Not all realists find the objection convincing, but, as noted above, many of those admitting the indeterminacy of phenomenal consciousness come close to illusionism in spirit. Schwitzgebel (2023) reacts to this objection by conceding that the positive inconceivability of indeterminate consciousness shows that our conceptualization of consciousness is inadequate, which entails at least a dose of weak illusionism.

This has implications for consciousness science. Major persisting debates, e.g., on the distribution question, might be dissolved and reframed. Interpreting the distribution question as bearing on quasi-phenomenal consciousness, our best comprehensive answers to the question probably have an indeterminate truth-value. Of course, many beings are determinately quasi-phenomenally conscious, and many are determinately not: think humans versus heaps of sand. However, in many difficult cases, which are precisely those about which consciousness scientists debate the most, it is likely that things are indeterminate, with no corresponding fact of the matter. No truth waits for us to be discovered there. Continuing the debates seems a waste of time. We are better off replacing the quasi-phenomenal distribution questions with alternative distribution questions concerning more specific properties of our quasi-phenomenal states, like *being a globally broadcast representation outputted by various modules working in parallel*. Those questions are more likely to admit of determinate answers.

To see how this could play out concretely, consider the debate on AI consciousness. Could non-biological AIs ever be conscious? Proponents of some varieties of functionalism say yes.[88] So-called biological naturalists say no.[89] The opposition is entrenched, and the debate is heated. My view suggests that there is probably no fact of the matter here, and that the opposition between the two camps should be dissolved. Interpreted as quasi-phenomenal consciousness, the best answer to the AI consciousness question is probably indeterminate. Non-biological AIs might have states like our quasi-phenomenal states in some respects, but not in others, without any of these respects having a privileged status here. This sort of indeterminacy-motivated dissolution and reframing could also apply to quasi-phenomenal extension and basis questions, even if this is less probable.

Suppose you concede that claims about the distribution, extension, and basis of consciousness, reinterpreted by illusionists as about quasi-phenomenal consciousness, are often indeterminate. Still, you could resist this reframing. You could contend that, when these claims are determinately true or false, they teach us much: they impart important knowledge to us—knowledge we give up if we only ask about other more specific properties of our quasi-phenomenal states. What seems important to know is whether a creature is *conscious*, not just whether it has globally broadcast

[88] See Butlin et al. (2023).
[89] See Seth (2025).

representations outputted by various modules working in parallel! If so, we should refuse the reframing I just championed, even if we concede indeterminacy. I address this concern in the next section.

Illusionism's challenge to the determinacy presupposition has long been noticed. Dennett denied that "there [is] a dividing line" between consciousness and unconscious features. He also denied that consciousness "is either present or absent."[90] Still, he leaned toward the view that only language-users have processes that "count as consciousness,"[91] suggesting that he viewed the *distribution* of consciousness among non-humans as determinate: if language is required for consciousness, non-human animals and pre-verbal infants, arguably, are determinately non-conscious. Dennett also thought the *extension* of consciousness is indeterminate in humans. He stressed that:

> one cannot, logically, make precise determinations about *when* consciousness of one item or another happens. (Dennett 1998, 105)

However, this indeterminacy is a mild one, following from sorite-like considerations of borderline cases. This sort of mild indeterminacy, linked to borderline cases, stems naturally from Dennett's "multiple drafts"[92] and "fame in the brain"[93] views of consciousness. It is not the hard-to-contain, radical indeterminacy for which I argued by considering the multiple properties of our own quasi-phenomenal states.

D. Indeterminacy and the illusion of phenomenal consciousness

Indeterminacy may also concern the *illusion* of phenomenality. Suppose, for instance, that Meta-Cartesianism correctly describes the illusion of phenomenality. If so, illusions of phenomenal experiences belong to various types. They are introspective representations (a), but also introspective representations tracking non-inferential belief-formation processes (b), as well as representations of primitive direct evidence (c), characterized as introspectively obvious (d), irreducible (e), and undetectable (f), etc. They are

[90] Dennett (1995a, 706).
[91] Dennett (1995a, 703).
[92] Dennett (1991).
[93] Dennett (2006).

also representations manipulated by an ERM (g). At a deeper level, which I did not speculate about, they involve neural activities belonging to variously determined neural types (α, β, γ, etc.).

The same argument as earlier can be built. Take a being—e.g., some future AI—with functionally similar processes (with a, b, c, d, e, f, g, etc.) but devoid of biological neurons (without α, β, γ, etc.). Or take an AI whose introspection tracks non-inferential belief-formation processes which it characterizes as primitive and irreducible (with a, b, c, e), but not as obvious and undetectable (without d and f), or without using an ERM (without g). Or take one characterizing its target as primitive, obvious, irreducible, and undetectable (with a, c, d, e, f) but that does not track non-inferential belief-formation processes (without b).

Some of these beings would be, in some respects, but not in others, under the same illusion as us. Plausibly, for some of those, whether they have the illusion of phenomenality is indeterminate. Therefore, the *potential* distribution of the illusion is probably indeterminate. "Potential" is key here: I will not argue for this, but it seems plausible that only possible minds, like future AIs, fall in the indeterminate zone. Non-human animals do not: more probably, they determinately lack the illusion of phenomenal consciousness.

As before, various responses are available here, such as the natural kind response, and counter-responses are similar to those previously given. Finally, there is also the possibility that the extension and basis of the illusion of phenomenality is indeterminate in us, but I do not think it is as likely, and I will not discuss it.

8.8. The informativity of consciousness claims

Finally, illusionism challenges the presupposition that claims about consciousness, notably when reinterpreted as about quasi-phenomenal consciousness, are *informative*.

Take the claim "Lobsters are quasi-phenomenally conscious." To simplify, suppose that it is determinately true. Given illusionism, this sort of claim will feature in responses to the distribution question. What do we learn when we learn this claim?

Before answering, consider what happens if we assume *realism*, and learn that lobsters are *phenomenally* conscious. Learning this, we learn that there is something it is like to be a lobster: lobsters phenomenally feel certain things.

Intuitively, this teaches us something substantive and important about the lobster: that it is an experiencing subject, that it has a certain sort of subjective point of view, that "the lights are on" or that "someone is home," to use frequently employed expressions. Claims of this sort seem highly informative, in the non-technical sense that they impart substantive and important knowledge to us. Not *all* realists think that claims about the distribution of phenomenal consciousness (or its basis and extension) are highly informative, but most do, at least implicitly.

Now, by contrast, what do we learn when, assuming illusionism, we learn that a lobster is *quasi*-phenomenally conscious?

First, imagine that consciousness science has not started. We have no scientific knowledge of quasi-phenomenality. The claim that lobsters are quasi-phenomenally conscious just means that they enter states of the same type as those tracked and mischaracterized by phenomenal introspection in us—nothing more. This teaches us very little about how lobsters' minds work and what they do. It just ascribes them a property grasped in a highly indirect and relational manner, in terms of how it is mischaracterized by us. The claim is not highly informative. Arguably, the same holds for claims about the basis or extension of quasi-phenomenality.

Thinking of sirens helps. If I learn that there are sirens—*genuine* human-like sirens—in the Bay of the Somme, in Northern France, I obtain substantive and important knowledge about the Bay's wildlife. The claim is highly informative because I have a certain substantive idea of what sirens are. By contrast, if I just learn that there are creatures of the same type as those that illusory sightings of sirens track (i.e., quasi-sirens), assuming I have no scientific knowledge about what these sightings track, this teaches me very little. For all I know, quasi-sirens could be dolphins, seals, fish, floating jellyfish, humans practicing artistic mermaiding, strangely shaped tree trunks, ship debris, clumps of algae, stranded inflatable dolls,[94] etc.

Still, one could retort that, *if* quasi-phenomenal consciousness happens to correspond to a single and specific natural kind, learning that lobsters are quasi-phenomenally conscious is informative, even if we know next to nothing about what this kind consists of. Indeed, correctly describing the lobster as instantiating this kind leads to various successful epistemic actions. We will group lobsters with humans and apes, look closely at what they have

[94] In 2016, a stranded inflatable sex doll discovered during a solar eclipse was mistaken for an "angel" or "spirit" (*bidadari*) by Indonesian fishermen. The doll was dressed, paraded, and photographed, until the police came to confiscate it (BBC News 2016).

in common cognitively, build hypotheses, draw inductions on this basis, etc. Since quasi-phenomenal consciousness is a natural kind, cutting nature at its joints, these epistemic actions will often be successful and will lead to important, unifying, and explanatory knowledge. The knowledge that lobsters are quasi-phenomenally conscious would be important in this (slightly indirect) sense, and the corresponding claim would be highly informative. Similarly, if, unbeknownst to me, quasi-sirens form a single specific natural kind (e.g., manatees), learning that there are quasi-sirens in the Bay is highly informative, as it allows for epistemic actions conducive to important knowledge.

I concede this. However, given illusionism, it is unlikely that quasi-phenomenal consciousness is a single specific natural kind. So, assuming that consciousness science has not started, a claim like "lobsters are quasi-phenomenally conscious," assumed here to be determinately true, is unlikely to be highly informative.

Now, suppose we know much, scientifically, about quasi-phenomenality. There are two possible cases.

First, suppose that quasi-phenomenality has in fact been shown to form a single specific natural kind F, which we know under a precise theoretical description ("F"). Learning that lobsters are quasi-phenomenally conscious tells us that they have F-states. In this case, the claim that lobsters are quasi-phenomenally conscious is highly informative, but not crucially so: it is mostly redundant with the claim that lobsters have F-states. At most, the first claim adds the idea that lobsters' F-states have human equivalents, which we misrepresent as phenomenal. Still, knowing that lobsters have F-states is the important thing here. Little is added by the claim that they are quasi-phenomenally conscious. The same holds for claims about the basis or extension of quasi-phenomenality.

Second, suppose that, as I argued is likely, quasi-phenomenality has been shown to fail to correspond to a single specific natural kind. Suppose also that we know about the underlying neural and functional features G_1, G_2, . . . , G_n, etc. (some of which might form specific natural kinds), whose presence and absence make it so that some creatures are determinately quasi-phenomenally conscious, some determinately *not* quasi-phenomenally conscious, and some neither. If so, learning that lobsters are quasi-phenomenally conscious is relatively *uninformative*. There are considerably more informative claims to learn in the vicinity, about features G_1, G_2, . . . , G_n. Compared to claims about G_1, G_2, . . . , G_n, claims about quasi-phenomenality are not simply redundant. They are potentially confusing, somewhat like

unspecified claims about creatures capable of memory once when we know that memory does not form a single, specific natural kind, but corresponds to various distinct specific kinds. Again, the same holds for claims about the basis or extension of quasi-phenomenality.

Let us summarize. If we have *no* scientific knowledge of quasi-phenomenality, claims about quasi-phenomenality are likely to fail to be highly informative. If we know a lot, scientifically, about quasi-phenomenality, claims about quasi-phenomenality are also likely to fail to be highly informative, since they are likely to be relatively uninformative. If we are lucky and quasi-phenomenal consciousness forms a single specific natural kind, claims about quasi-phenomenality *could be* highly informative; still, once we know a lot scientifically, they are not *crucially* highly informative.

I focused on two extreme cases: complete ignorance and complete scientific knowledge. Intermediate cases are analyzable as a mix of those: in those cases, too, claims about quasi-phenomenality are likely to fail to be highly informative.

This challenge to the informativity presupposition has one important immediate implication. Earlier, I explained that, as we abandon the *determinacy* presupposition about quasi-phenomenal consciousness, for instance, when trying to answer the distribution question, corresponding debates must be dissolved and reframed. The distribution question should transform into a set of questions about the distribution of more specific properties, like *being a globally broadcast representation outputted by various modules working in parallel*. One worry was that, if we stop asking the original consciousness questions, even reinterpreted as about quasi-phenomenality, we lose potentially important knowledge. We now see why this should not worry us too much: claims about quasi-phenomenal consciousness are unlikely to be highly informative.

Does this challenge to the informativity presupposition similarly concern claims about the *illusion* of phenomenal consciousness? Maybe not to the same extent. These claims might be more informative as such than those about quasi-phenomenality. Prior to scientific inquiry, we only know of quasi-phenomenality that it is tracked and mischaracterized as phenomenal, but we know more about the illusion itself. We might know a priori, with conceptual analysis and thought experiment, that an illusion of phenomenality is an illusion of introspectively obvious, irreducible, and undetectable states (but see the earlier concern about the indeterminacy). This could render

claims about the illusion of phenomenality informative. Again, the siren analogy helps: prior to inquiry, we might know more about illusory sightings of sirens (e.g., they represent human-like marine creatures with caudal fins) than about what they track.

Let us take stock. For a realist, claims about phenomenality seem highly informative. For an illusionist, claims about quasi-phenomenality are unlikely to be such. If we are scientifically ignorant of quasi-phenomenality, these claims are not highly informative, except if quasi-phenomenality corresponds to a specific natural kind, which I argued is unlikely. If we know much scientifically about quasi-phenomenality, then, if quasi-phenomenality corresponds to a specific natural kind, these claims are informative, but not crucially so. If, as is likely, quasi-phenomenality does not correspond to a specific natural kind, these claims are simply confusing. Setting aside quasi-phenomenal questions is unlikely to result in knowledge loss.

Quietists will stress that illusionism is not needed to defend this idea. Some non-illusionists, like Peter Carruthers, also think that claims about phenomenality are not highly informative.[95] However, illusionism implies this stance naturally, and non-illusionists taking the same stance, like Carruthers, come close to illusionism in spirit. Finally, one might object that even if claims about quasi-phenomenality are not highly informative *scientifically*, they could be ethically or epistemologically informative—for instance, if quasi-phenomenality is important for justification, well-being, or moral status. This last issue is discussed in Chapters 9 and 10.

8.9. Siren hunting: An illusionist instruction manual

Against the *radical* view, illusionism does not make consciousness science moot. Standard consciousness science questions bearing on phenomenal consciousness find heirs in an illusionist framework: questions about quasi-phenomenal consciousness and the illusion of phenomenal consciousness. Still, against the quietist view, illusionism has substantive implications regarding how consciousness science must be conducted.

Let us summarize these implications, starting with the most straightforward ones. Assuming illusionism, the putative capacity to solve the hard problem should not guide our assessment of scientific theories of

[95] Carruthers (2020).

consciousness. Scientists should stop fearing that consciousness might remain inaccessible to science. They should be wary of relying on introspective axioms regarding the nature of consciousness. The illusion problem and the meta-problem of consciousness should become central to consciousness science.

Other deeper implications are less obvious. They stem from illusionism's challenge to three widely shared implicit presuppositions of mainstream consciousness science. First, illusionism challenges the *unity* presupposition. Illusionist researchers doing consciousness science face at least two sets of questions about two objects, not one. This forces researchers to disambiguate their claims. Conciliatory dissolutions of theoretical oppositions become attractive. Second, illusionism challenges the determinacy presupposition. Answers to the *quasi-phenomenal distribution question* are likely to have an indeterminate truth-value. This indeterminacy is also likely to concern the quasi-phenomenal extension and basis questions. Less probably, it could apply to questions about the illusion of phenomenality. This indeterminacy is more probable, given illusionism, than the indeterminacy of phenomenal consciousness given realism. Corresponding debates might thus have to be dissolved and reframed as bearing on more specific properties of our mental states uncovered by science. Third, illusionism challenges the informativity presupposition. Claims about quasi-phenomenality, contrary to claims about phenomenality for a realist, are likely to fail to be highly informative. Ceasing to pursue answers to quasi-phenomenal questions is unlikely to cause much knowledge loss.

What general lesson emerges from all this? For illusionists, research on quasi-phenomenal consciousness—the most natural heir of phenomenal consciousness research given illusionism—should be relatively deprioritized compared to research on phenomenal consciousness for realists. It is less likely to deliver determinately true and highly informative claims than research on phenomenal consciousness for a realist. We should devote relatively fewer resources to it.

Cognitive science should pursue research on the functioning of human, animal, and artificial minds. We should better understand how and why beings introspect in various ways.[96] We should solve the illusion problem and the meta-problem of consciousness. We will then gain a better

[96] See Kammerer and Frankish (2023b) for a framework that aims at studying different possible forms of introspection across creatures, including phenomenal introspection.

understanding of the illusion of phenomenal consciousness and what it tracks: quasi-phenomenal consciousness. However, this knowledge of quasi-phenomenality should come as a bonus. It should not be our primary objective, since it is unlikely to constitute important knowledge. Answering quasi-phenomenal questions should not be a lodestar of consciousness science: as I already stressed elsewhere, it should be "an afterthought, not a guideline."[97]

We will thus also better assess whether quasi-phenomenality corresponds to a single specific natural kind. Supposing it does, we should positively characterize this kind. If GNWT describes it, for instance, we could characterize it as conscious access: we could *identify* quasi-phenomenal consciousness and conscious access. Claims about the distribution, extension, and basis of quasi-phenomenal consciousness would thus become highly informative, but not crucially so: they would be mostly redundant with claims about conscious access. Until we have discovered this putative single specific kind, we should keep using a cumbersome expression like "quasi-phenomenal consciousness"—or, maybe, "phenomenal*ish* consciousness," "conscious*ish* experiences," etc. Our wording must make it clear that what we study is *not* the entity depicted by introspection, of which we have a substantive idea. It is another entity, deeply mischaracterized by introspection, which we grasp only indirectly and in terms of how it is mischaracterized.

If, as I argued is likely, quasi-phenomenal consciousness is not ultimately found to correspond to a single specific kind, but, say, to distinguishable kinds which can each be studied for their own sake, consciousness scientists should simply stop using the concept of quasi-phenomenality in most scientific endeavors. They should replace it with concepts capturing the relevant distinguishable kinds in a context-sensitive manner. Scientific claims about quasi-phenomenal consciousness should be replaced with conceptually heterogeneous claims about these natural kinds, thereby carving up minds differently.

Again, the siren analogy helps. Once we know genuine human-like sirens do not exist, open questions remain about illusory sightings of sirens and what they track. Still, focusing on these questions seems a mistake. Marine biologists should study marine fauna without prioritizing the identification of quasi-sirens. Learning more about manatees, they could come to realize that they are, in fact, quasi-sirens: creatures reliably tracked and

[97] Birch et al. (2022, 17).

mischaracterized by past sailors' sightings of sirens. While interesting, establishing this should not be a primary aim of marine biology. We should *not* conclude that this field failed in a deep sense if it fails to establish whether sailors like Columbus were, in fact, facing manatees or dolphins when they thought they saw sirens. Learning that manatees are quasi-sirens is *not* highly informative. It does not tell us much about what manatees are. Establishing the nature of their life cycle, behavior, metabolism, cognitive capacities, etc., is much more important than establishing that they are quasi-sirens. Similarly, knowing that some mental states are tracked and mischaracterized as phenomenal by introspection does not tell us much about what they are. Besides, there might be no fact of the matter when it comes to answering the following questions: Are manatees the 'real' quasi-sirens? Do they share this title with dolphins, seals, fish, floating jellyfish, sirenidae (such as the *Sirenia lacertina* found by Linnaeus), or clumps of algae? Arguably, there is no truth to be discovered there. Similarly, there might be no truth to be discovered regarding the real nature of quasi-phenomenal consciousness. Let us avoid making research on quasi-phenomenal consciousness a priority.

This sort of consideration percolates through consciousness science and impacts it in various subtle ways. I cannot list them all, but I will focus on two examples. One concerns so-called theory-neutral approaches to the distribution question. The other concerns the natural kind approach to consciousness.

First,[98] take the theory-neutral approach (or analogy approach) to the distribution question.[99] Researchers taking this approach first establish *indicators* of phenomenal consciousness: behavioral or cognitive features that specifically correlate with phenomenal experiences in humans. When these features are found in other creatures, researchers conclude that they are more likely, ceteris paribus, to be phenomenally conscious. Specific features, like the capacity to engage in *trace-conditioning* (a form of psychological conditioning where the two associated stimuli are separated by a temporal interval), become the focus of study,[100] as they supposedly indicate phenomenal experiences, for example since there is tentative evidence that humans are only capable of trace-conditioning about consciously experienced stimuli.[101]

[98] The reasoning deployed in this paragraph and the next is developed at length in Dung and Kammerer (2025).
[99] For an example, see Tye (2017). See Birch (2022b) for the label.
[100] Allen (2013); Birch et al. (2020); Birch (2022b).
[101] R. Clark and Squire (1998; 1999)

Given illusionism, features like the capacity to engage in trace-conditioning cannot indicate *phenomenal* consciousness. They may indicate quasi-phenomenal consciousness, but since illusionists should de-prioritize research on quasi-phenomenal consciousness, they should also, to some extent, de-prioritize research on these features. This does not imply that these features are unimportant, only that they plausibly become relatively less important to study if our consciousness-based interest disappears.

Now, turn to the natural kind approach (or iterative approach), presented earlier in the chapter. This approach also starts with indicators of consciousness, but instead of directly using the distribution of indicator properties to infer the distribution of consciousness, it enters an iterative process of hypothesizing, testing, and refining the indicators, hopefully zeroing in on the natural kind underlying consciousness.

I argued that illusionism makes it relatively less likely that quasi-phenomenal consciousness forms a natural kind, compared to what happens for phenomenal consciousness given realism. So, searching for a natural kind underlying consciousness becomes a relatively less promising bet. This certainly does not mean that an illusionist version of the natural kind approach is guaranteed to fail: I conceded that quasi-phenomenal consciousness could still turn out to correspond to a natural kind.

Finally, the siren analogy illuminates consciousness science in a last manner. Phenomenal consciousness is akin to *mythical* sirens. Mythical sirens had luring and tempting voices, and legendary seafarers could not resist joining them—and drowned. Phenomenal consciousness, too, is tempting. We hardly resist believing in its existence. It seems mysterious and significant, and it easily fascinates us. We tend to see it as a unitary and determinate phenomenon, whose knowledge is highly important. Fascinated researchers might then wrongly prioritize answering the consciousness questions. They will continue to feel that they should study conscious experience, or phenomenal consciousness, or what it's like to be someone, rather than discuss more boring, non-fascinating phenomena, like informational access or higher-order representations. They will keep asking: are lobsters, octopuses, or LLMs *really* conscious? The temptation will always keep reemerging. We must guard ourselves against it.

In this respect, strong illusionism has a virtue that verbal weak illusionism lacks. Its use of terms makes the difference between the fascinating phenomenon (phenomenal consciousness) and the boring one (quasi-phenomenal consciousness) unmistakable. Strong illusionists make it clear that we should

only scientifically study the boring phenomenon, since the fascinating one is not real. Verbal weak illusionists muddy the water and make the temptation harder to resist. I am fairly pessimistic: I doubt that a majority of consciousness scientists will ever adopt this perspective. Studying the fascinating thing might simply be too tempting. Still, it might be worth trying to convince them otherwise.

8.10. Illusionism and paradigm shift

At this stage, some readers will insist that there can be no such thing as an illusionist science of consciousness. Denying phenomenal consciousness means denying the explanandum of consciousness science. What remains after this is not consciousness science. Similarly, someone who denies the existence of planets and stars cannot be an astronomer.

This reproach, and the opposition between realist and illusionist consciousness science, areclarified by the concept of paradigm shift. Some episodes of the history of science have been interpreted as competitions between scientific paradigms. "Paradigm," here, refers to a set of key theories, data, methods, values, and metaphysical assumptions constituting a disciplinary matrix, on the basis of which research is pursued, and which is not typically interrogated in research.[102] Competing paradigms are *incommensurable*. For instance, each paradigm comes with its own methods, evaluation criteria, and basic data. Therefore, supporters of competing paradigms often agree on too little to reach consensus by rational means when it comes to paradigm choice. When incommensurable paradigms compete, each is often preferable by its own lights.

Defenders of a dominant existing paradigm can frequently point to basic, obvious data that the other paradigm cannot explain. Aristotelian geocentrists opposed Galilean heliocentrists by appealing to the *obvious fact that the Earth does not move*. They used the 'tower argument', observed that a stone falls from a tower perpendicularly to the surface of the Earth, and insisted that it shows that the Earth does not move—as it would indeed if the Aristotelian conception of motion as absolute were correct. According to Paul Feyerabend:

[102] Kuhn (1962).

> the idea of the motion of the earth [was] outlandish, absurd and obviously false, to mention only some of the expressions which were frequently used at the time.[103]

Heliocentrists, in response, devised persuasive rhetoric to convince the undecided to discard their opponents' data. They claimed that the only data we have regards the *apparent* immobility of the Earth (relative to us, not to the Sun) and the *apparent* straight and perpendicular motion of the stone (relative to the surface of the Earth, not to the Sun). This data is unproblematic for heliocentrism. The 'data' of Geocentrists is only undeniable if one implicitly accepts false views of motion.

Plausibly, illusionism and realism are incommensurable paradigms in consciousness science. Realism starts with phenomenal consciousness as its basic, obvious explanandum. Illusionism aims at explaining its non-phenomenal apparent existence.

One limit of the analogy with geocentrism and heliocentrism is that, arguably, we are much more reluctant to discard data regarding the existence of phenomenal experiences than we were to reject data about the Earth's immobility. Meta-Cartesianism accounts for this by explaining why we cannot intuitively make sense of the idea that phenomenal experiences appear to exist without existing. The reconceptualization of data involved in past paradigm shifts appealed to our intuitive sense of what constitutes basic, undeniable data (phenomenal experiences), an intuitive sense stemming from our innate apparatus representing evidence. Heliocentrists could thus claim that the Earth moves even if we perceptually experience it as motionless. This reconceptualization of data was counterintuitive, but it was still graspable by all as it appealed to an intuitive bedrock of data: our perceptual experiences of a still Earth.

By contrast, the predicament of illusionists is unique. They reconceptualize data without appealing to the same intuitive bedrock of data, the experiential bedrock being precisely what is in discussion. While paradigm competitions involve disagreement over data, the conflict between illusionists and realists also concerns what are legitimate ways of reconceptualizing our data, and the illusionist's approach is also counterintuitive at this level.

When D.C. Williams, in the 1930s, criticized eliminativist interpretations of behaviorism, he claimed that consciousness is "the only thing wholly

[103] Feyerabend (1993, 61). Feyerabend, himself a proponent of eliminativism, draws the parallel between heliocentrism and eliminativism (Feyerabend 1993, 123–24).

'empirical' and immediately 'observable.'"[104] Citing Wolfgang Köhler, Edwin Boring, and Max Planck, he insisted that it forms the basic data of *all* of science.[105]

> [eliminativism] entails the most searching scientific revolution ever conceived, not merely in psychology but in all human concept-systems, and all logical and scientific methodology. Such revision, although not impossible, is greater than any attempted by a Plato, a Darwin, or an Einstein.[106]

Pace Williams, the illusionist revolution is, in most respects, much more modest than those; it requires little conceptual innovation. However, in one respect, Williams was right: the illusionist paradigm uniquely puts into discussion our most intuitive way of apprehending our bedrock of data, something that past scientific revolutions, arguably, had left untouched.

Historically, the outcome of paradigm competitions was not decided by appealing to commonly accepted methods. Winning paradigms were those that offered the best promises of success.[107] I tried to show that, while realist consciousness science faces fundamental problems, illusionism presents a viable alternative. It carries convincing promises of success. Hopefully, this will help convince the undecided.

* * *

This chapter argued that illusionism has serious, non-simplistic, and non-trivial implications for consciousness science. The next chapter examines the ethical implications of illusionism.

[104] Williams (1934, 465).
[105] Williams (1934, 466–67).
[106] Williams (1934, 468).
[107] Kuhn (1962, chap. 3).

9

Losing Your Religion

Ethics After Illusionism

> *The conscious world of feelings, intentions, and evaluations, is only a small parcel in the immensity and multiplicity of cooperating and conflicting processes that the life of any organism contains. We have no right to consider this bit of consciousness as the purpose, as the "why?" for the entire phenomenon of life*
>
> —Friedrich Nietzsche

Phenomenal consciousness seems ethically significant. Does denying its existence, as illusionism does, have ethical implications? If so, which ones? This chapter tackles these questions.

First, I motivate the exploration of illusionism's ethical implications. I sketch a religious analogy to guide my inquiry. I then focus on our intuitive ethical picture, which gives an important ethical role to phenomenality. I show that, compared to this intuitive picture, illusionism has revisionary implications regarding the grounds of well-being and moral status, but not necessarily for their distribution. Potential revisionary distributional consequences depend on the positive non-sentientist ethical view that illusionists endorse. I sketch four non-sentientist ethical views, assess them, and present their consequences. I close the chapter with concluding remarks.

9.1. Motivating the exploration

Does illusionism have ethical implications? There are at least three reasons to ask the question.

First, if we have non-negligible credence in illusionism, we must answer it, so that we can update our ethical beliefs and practices appropriately.

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0010

Second, knowing illusionism's ethical implications helps to assess the view, since these implications motivate objections, typically appealing to *ethical data* regarding supposedly manifest ethical truths (e.g., "pain is bad for us" or "humans have moral status"). One objection aims at directly refuting illusionism. It contends that illusionism entails the negation of some manifest ethical truth and can thus be ruled out directly. A more modest objection concedes that illusionism alone does not contradict manifest ethical truths, but does so when conjoined with plausible premises. If so, illusionism struggles to account for ethical data, and we can draw an inference to the best explanation of ethical data against it. In the first objection, ethical data is treated as direct data against illusionism; in the second objection, it is treated as indirect data.

Third, even if we bracket the existence of ethical data or doubt that ethical premises can provide epistemic reasons to reject metaphysical views,[1] illusionism's ethical implications can still give pragmatic reasons to reject the view. If illusionism, alone or with plausible premises, implies that intuitive ethical beliefs are false, illusionists will tend to reject these beliefs and might engage in behavior we find undesirable. This could give us pragmatic reasons to reject the view, to hinder its adoption. In this vein, Galen Strawson wrote:

> If [Dennett's illusionism is] right, no one has ever really suffered, in spite of agonizing diseases, mental illness, murder, rape, famine, slavery, bereavement, torture, and genocide. And no one has ever caused anyone else pain [. . .] We must hope that [this idea] doesn't spread outside the academy, or convince some future information technologist or roboticist who has great power over our lives.[2]

Katalin Balog expressed the same concern:

> [B]eing dismissive towards conscious experience makes one unable to fully account for the moral worth of sentient beings. It predisposes anyone believing in it to see humans in the same way most of us think about robots: unfeeling, meaningless (meat)-machines [. . .].

[1] Impagnatiello (forthcoming).
[2] Strawson (2018).

> A belief in illusionism is not merely wrong, it leads to morally detrimental consequences.[3]

Therefore, we must assess illusionism's ethical implications. My assessment draws on three presuppositions.

First, we share a certain *intuitive ethical picture*—"we" meaning at least most contemporary philosophers of mind. We tend to see phenomenality as real and as playing a role for well-being and moral status. We agree broadly on many aspects of the distribution of well-being and moral status: humans can have very low or very high well-being, some animals (cows, bats, falcons, etc.) probably also have a degree of well-being, but plants, hair driers, and rocks probably do not; humans have high moral status, complex animals probably have some moral status, but plants, hair driers, and rocks probably do not. About certain difficult cases, our intuitive picture hesitates: Do bees, lobsters, worms, ants, sophisticated AIs, or cerebral organoids matter morally? Are they capable of welfare?

One advantage of assuming the existence of this intuitive starting point is that we can put illusionism's implications in terms of revisions: Does illusionism imply or suggest revising this picture?

Second, for the sake of the argument, I grant that some intuitive ethical claims are manifest. They possess strong pre-theoretical justification, and can legitimately count as ethical data, at least for now. Claims like "pain is bad for us" or "humans have moral status" might qualify.

Third, I assume *ethical cognitivism*: claims about prudential value and moral status express beliefs, rather than non-truth-evaluable mental states like emotions or intentions. Ethical cognitivism is plausible, and it makes the discussion easier. Still, it is not required to explore illusionism's ethical implications. If ethical claims express emotions or intentions, we can still ask whether adopting illusionism would or should change these emotions or intentions.

9.2. Godless Duties

The last chapter used a siren analogy. This one uses a religious analogy.

Until recently, the West was overwhelmingly Christian. Belief in God was treated as a given. Avowed atheism was virtually non-existent before the

[3] Balog (2023, 15–16).

18th century, and only became the view of a substantive minority during the 19th century. Before this, it seemed obvious to most that, without God, there were no *moral duties*: coherent atheists *had* to deny such duties. A 'virtuous atheist' seemed a paradox, whose possibility had to be argued—famously, by Pierre Bayle.[4] Even critics of traditional religion, like Voltaire, thought that atheists would obviously be immoral:

> In what does a society of atheists appear impossible? It is that one judges that men who had no check could never live together; that laws can do nothing against secret crimes; that a revengeful God who punishes in this world or the other the wicked who have escaped human justice is necessary.[5]

The idea remained long influential. Dostoevsky put it in the mouth of Ivan Karamazov:

> Without God and the future life? It means everything is permitted now, one can do anything?[6]

We now live in the 21st century. Atheism grows throughout the world. In the West, it is the majority view among the youngest and most educated parts of the population. Voltaire's fears seem dated: society keeps functioning, and it would do so even if everyone stopped believing in God. Most atheists are reluctant to commit even "secret crimes." Many sincerely think that certain things are morally right or wrong.[7]

The religious analogy will help us assess illusionism's ethical implications. Before this, however, we must characterize our intuitive ethical picture.

9.3. The intuitive character of sentientism

Our intuitive ethical picture relies on *phenomenal realism* and *sentientism*. It views the phenomenal as real and as playing a role for prudential value and moral status.

[4] Bayle (1699).
[5] Voltaire ([1764] 1924, entry "Atheism").
[6] Dostoevsky ([1880] 1992).
[7] For some empirical data on atheists' moral values, see Ståhl (2021).

A. Phenomenality and prudential value

Consider the idea that the phenomenal plays *some* role for prudential value:

> *Phenomenal Value*: Some standard phenomenal states ground notable value.

"Value" here, and always in this chapter, means prudential value, value-for-a-subject—directly contributing to the subject's well-being. "Notable value" means non-negligible value. *Standard* phenomenal states are states that phenomenal realists think frequently occur, like perceptual experiences, pain, pleasure, etc. "Ground" means *fully ground*—it corresponds to the *fully in virtue of* relation. It is a shortcut for "ground or would ground": *Phenomenal Value* can be true even if there are no phenomenal states, as long as, *if* certain standard ones occurred, they would ground value. Illusionists can coherently accept *Phenomenal Value*.

Phenomenal Value is extremely intuitive. Consider this:

> *Pure Suffering*: Piotr suffers from a unique condition. Every night, after some sleep, he wakes up. For one hour, he experiences intense phenomenal pain without engaging in any observable behavior, before going back to sleep. In the morning, he retains no memory of the painful episode. The episode does not change any of his psychological dispositions.[8]

Intuitively, what happens then is *bad* for Piotr because of how it phenomenally feels. Since phenomenal pain is standard, *Pure Suffering* lends support to *Phenomenal Value*.

One can defend a stronger claim:

> *Distinctive Phenomenal Value:* Some standard phenomenal states ground notable value, and they ground distinctive value compared to other states.

Distinctive Phenomenal Value conjoins *Phenomenal Value* with a comparative claim. "Other states" refers to non-phenomenal states otherwise close to phenomenal states, e.g., in terms of complexity, or their relation to behavior or other cognitive processes. "Other states" should be read as

[8] Adapted from Kammerer (2020, 899).

implicitly restricted to positively conceivable states, leaving space for non-phenomenal states we cannot conceive of grounding the same value as phenomenal states. The mental states of zombies are an ideal example of such other states: they are positively conceivable, non-phenomenal, but otherwise as close to phenomenal states as possible. Distinctive value is value quantitatively or qualitatively much more significant than other values. Again, "ground" is a shortcut for "ground or would ground." Illusionists can coherently accept the claim.

Distinctive Phenomenal Value is intuitive. Consider this:

> *Zombification*: Alyona and Lizaveta are two phenomenally conscious sisters. Alyona is healthy, successful, loved, and lives a happy life. Lizaveta is solitary, unsuccessful, and suffers from excruciating chronic diseases; she often contemplates suicide and only abstains because her death would pain her mother. One day, a demon zombifies both of them. They lose all phenomenal experiences while retaining their physical properties.[9]

Intuitively, pre-zombification, Alyona's life contains a lot of positive value, while Lizaveta's life contains a lot of negative value. Intuitively, zombification changes things. Alyona's life does not seem as good for her. Lizaveta's life does not seem as bad. The two sisters' lives seem prudentially neutral, or at least closer to neutral than before. By stipulation, they only lost phenomenal experiences. This suggests that phenomenal states ground distinctive value compared to other states. If one already accepts *Phenomenal Value*, this seems enough to support *Distinctive Phenomenal Value*.

Some readers might be perplexed. Illusionists, like myself, think the phenomenal is not real. Can we intuit its ethical role? Others, like realist physicalists, think zombies are impossible. Can they intuit their well-being? Here, I assume that both phenomenally conscious beings and zombies are conceivable, and that we can have intuitions about conceivable situations even if they are impossible. I intuit that living in Heaven would be good for me, or that angels would have moral status, even if I doubt that these things are possible. Here, I follow William Seager:

> [Q]uestions of value assignment are independent of the kind of possibility which the thought experiment depends upon or exploits. We can assign, if

[9] See Siewert (1998; 2014) for zombification arguments.

only hypothetically, value to things that are physically impossible or even metaphysically impossible.[10]

Finally, here is an even stronger claim:

Exclusive Phenomenal Value: Some standard phenomenal states ground notable value, and nothing non-phenomenal grounds value.

Exclusive Phenomenal Value conjoins *Phenomenal Value* with the claim that nothing non-phenomenal grounds value. Again, the claim should be read as implicitly restricted to positively conceivable properties, leaving space for non-phenomenal value-grounding properties we cannot conceive of. *Exclusive Phenomenal Value* entails *Distinctive Phenomenal Value*, but not conversely.

Exclusive Phenomenal Value is somewhat intuitive. Consider this:

Zombie World: A zombie world is physically and functionally identical to the world the phenomenal realists thinks we inhabit, but contains no phenomenal experiences.

Some find it intuitive that, in *Zombie World*, no one has a good or a bad life. For someone who already accepts *Phenomenal Value*, this seems enough to support *Exclusive Phenomenal Value*.

B. Sentience and moral status

Now, turn to moral status. First, consider a modest claim:

Sentience Criterion: Sentience grounds moral status.

Sentience Criterion means that sentient creatures, in virtue of their sentience—their capacity to have valenced phenomenal experiences—have, or would have, certain moral status. It is extremely intuitive. It follows if we admit that valenced phenomenal states ground prudential value—something hard to deny for those who accept *Phenomenal Value*—and that

[10] Seager (2001, 3).

welfare subjectivity is sufficient for moral status—a thesis I granted is highly intuitive.

Sentience Criterion is also directly intuitive. Consider this:

> *Mystery Trolley*: A train runs on a track. You can stop it at a small cost. You know that an entity E is on the track and will be destroyed by the impact. You first know nothing about E. You then learn that E is sentient.

Intuitively, learning that E is sentient gives you a pro tanto moral duty to stop the train, suggesting that E has moral status in virtue of its sentience. This makes *Sentience Criterion* highly plausible.

One can go beyond *Sentience Criterion*:

> *Distinctive Sentience Criterion*: Sentience grounds moral status, and it grounds a distinctive sort of moral status compared to other capacities.

This combines *Sentience Criterion* with a comparative claim. "Other capacities" are capacities other than sentience—which, for the sake of the argument, do not mobilize anything phenomenal—but otherwise close to sentience, e.g., in terms of complexity, or relation to behavior and other cognitive processes. Again, there is an implicit restriction to the capacities we can positively conceive of, and again, the capacities of zombies are an ideal example. A *distinctive* moral status corresponds to a much higher degree of deserved moral consideration.

If we think that valenced phenomenal states ground distinctive prudential value—an idea that fits well with *Distinctive Phenomenal Value*—and that the degree of well-being a creature can attain determines its degree of moral status, *Distinctive Sentience Criterion* follows. However, the claim is also directly intuitive. Consider this:

> *Zombie vs Sentient Humans Trolley*: Sentient humans walk on a train track. Their zombie twins walk on another track. Two trains are running—one on each track—and are about to destroy all these creatures. Only one train can be stopped—at a small cost.

Should you stop one of these trains? If so, which one? Suppose both groups contain one individual. Many intuit that they should save the sentient human. Some might intuit that they should save one human even if there are

two, five, twenty (or more) zombies on the other track: sentient humans have much higher moral status. *Distinctive Sentience Criterion* seems intuitive.

Finally, we could go even further.

Exclusive Sentience Criterion: Sentience grounds moral status, and only sentience grounds moral status.

Again, there is an implicit restriction to capacities we can positively conceive of. The claim entails *Distinctive Sentience Criterion*, but not conversely. *Exclusive Sentience Criterion* follows if we think that valenced phenomenal experiences are necessary and sufficient to ground welfare—a view which fits well with *Exclusive Phenomenal Value*—and that welfare subjectivity is necessary and sufficient for moral status. It is also directly intuitive to some extent. Consider this thought experiment:

> *Zombie Trolley*: Zombies walk on a train track. A train runs on the track and is about to kill the zombies. You can stop it (and thus save the zombies) at a small cost.

Should you stop the train? Some will have the impression that letting zombies die is acceptable, suggesting that non-phenomenal capacities do not ground moral status. Putting aside certain considerations regarding broad sentience, for someone who already accepts *Sentience Criterion*, this seems enough to support *Exclusive Sentience Criterion*.

C. Varieties of sentientism

With these six intuitive claims, we can describe three varieties of ethical sentientism: weak sentientism, distinctive sentientism, and exclusive sentientism:

Weak sentientism: the phenomenal plays some ethical role (i.e., for well-being and moral status).
Distinctive sentientism: the phenomenal plays a distinctive ethical role.
Exclusive sentientism: the phenomenal plays an exclusive ethical role.

Weak sentientism conjoins *Phenomenal Value* and *Sentience Criterion*. Distinctive sentientism conjoins *Distinctive Phenomenal Value* and *Distinctive Sentience Criterion*. Exclusive sentientism conjoins *Exclusive*

Phenomenal Value and *Exclusive Sentience Criterion*. All three forms of sentientism are intuitive to some extent. All are technically compatible with illusionism.

Does this capture all views reasonably called "sentientist"? No. Suppose you think sentience is necessary but not sufficient for moral status. Or, suppose that, although it is neither necessary nor sufficient for moral status, it significantly *increases* moral status. If so, your view contradicts even *Sentience Criterion*, but it seems to deserve the label "sentientism." Still, the tripartition above captures many popular options and allows us to discuss illusionism's implications.

9.4. The wide acceptance of sentientism

Sentientism is widely accepted. First, many views of well-being imply that some phenomenal experiences are necessary and sufficient for value, which entails *Exclusive Phenomenal Value*. It is the case of hedonism, on which nothing has value but pleasure and pain.[11] At least, this is the case if one understands pleasure and pain *phenomenally*. Hedonists who understand pleasure otherwise, notably in terms of attitudes,[12] might still avoid *Exclusive Phenomenal Value*.

Moreover, many reject hedonism strictly speaking, but accept closely related experientialist views of welfare, on which nothing has value but certain phenomenal experiences.[13] This also entails *Exclusive Phenomenal Value*, or close-enough claims.

Beyond hedonism and related views, influential views of well-being include desire-satisfaction theories, on which what contributes to well-being is the satisfaction of one's desires, and objective-list theories, on which various objective goods beyond pleasure and desire-satisfaction contribute to welfare.[14] In many versions of these views, the phenomenal is important for

[11] The view had proponents in Ancient times (Epicureans, Cyrenaics, Democriteans, etc.) and in the 18th and 19th centuries (La Mettrie, Bentham, Mill). With some complexities and qualifications, it still has defenders (Crisp 2006, 103–9; Sprigge 1988, chap. 5; Tännsjö 1998, 84; for an overview, see Moore 2013).

[12] Feldman (2002). Traditional hedonism is not always clear on whether pleasure and pain are feelings, attitudes, or something else. Early Western hedonists like Bentham and Mill were probably leaning toward feelings (Moore 2013). Sidgwick initiated attitudinal views (Sidgwick 1907).

[13] Views of the sort abound (Brentano 1952; Kriegel 2018, chap. 9; 2019; A. Lee 2025; Seager 2001; van der Deijl 2019, 2021).

[14] Crisp (2021).

value. For instance, one influential version of the desire-satisfaction view states that what matters for well-being is a kind of *subjective* satisfaction of one's desires, which is no other than pleasure. This version of the approach is also a version of hedonism.[15] If one thinks of pleasure as essentially phenomenal,[16] this entails *Exclusive Phenomenal Value*. Most objective-list theorists also admit that some goods of the list are certain phenomenal experiences. This entails *Phenomenal Value*, and, in some versions, *Distinctive Phenomenal Value*.

Phenomenal Value is almost universally accepted.[17] Some downplay the importance of phenomenal consciousness for prudential value[18]: they insist that non-phenomenal states *also ground* significant value, thus rejecting *Exclusive Phenomenal Value* and maybe *Distinctive Phenomenal Value*, but not *Phenomenal Value*. The most cautious examinations grant that something like *Phenomenal Value* is "at least defensible":[19] phenomenality plays at least "a minor role" for well-being.[20] There are only a few exceptions.[21]

When it comes to moral status, *Distinctive Sentience Criterion*, the idea that sentient beings have, in virtue of their sentience, distinctive moral status, is widely accepted.[22] This idea motivated the 2012 "Cambridge Declaration on Consciousness" and the 2024 "New York Declaration on Animal Consciousness."[23] It seems assumed in the 2007 EU Lisbon Treaty (Article 5b),[24] and in the Global Animal Welfare Strategy of the World Organization for Animal Health.[25]

Some have argued that the capacity to have phenomenal experiences in general, and not just valenced ones, is necessary and sufficient for moral

[15] Heathwood (2006).
[16] Which might not be Heathwood's own view (Heathwood 2011, 91–93).
[17] Some have claimed that phenomenality in general grounds value. This entails (but is not needed) for *Phenomenal Value* (Glannon 2016; Nagel 1979, 79; for a discussion, see A. Lee 2018).
[18] Carruthers (1999); G. Lee (2014); Levy (2014).
[19] Levy and Shepherd (2020, 657).
[20] Levy (2024, 21). Shepherd defended a variety of sentientism (Shepherd 2018), but lately became more skeptical (Shepherd 2022). Levy is among the most skeptical of sentientism in recent philosophy.
[21] One might be Delon (2024), who doubts, from a Nietzschean perspective, that suffering grounds intrinsic negative value, which opens the door to denying *Phenomenal Value*.
[22] There are many defenders of *Distinctive Sentience Criterion* (Birch, 2022a; 2024; Bradford (2023); Dung 2024; Korsgaard 2018; A. Lee 2025; forthcoming; Lin 2021; Mellor 2019; Nussbaum 2024; Regan 2004; Schwitzgebel and Garza 2015; Sebo 2018; Shepherd 2016; 2018; Singer 1993; van der Deijl 2021). A number of them even accept *Exclusive Sentience Criterion*.
[23] Low et al. (2012); Andrews et al. (2024).
[24] European Union (2007).
[25] Dawkins (2021, 87).

status: *broad* sentience, rather than narrow sentience, plays an exclusive role for moral status.[26] This rules out *Exclusive Sentience Criterion*, but not necessarily *Distinctive Sentience Criterion*, and the resulting view is anyway close in spirit.

Some contend that capacities other than sentience ground moral status. This threatens *Exclusive Sentience Criterion*, but not necessarily *Distinctive Sentience Criterion*. Some think that *agency*—roughly, the capacity to act in goal-directed manners—grounds moral status, while also endorsing something like *Distinctive Sentience Criterion*.[27] Some resulting views are compatible with the idea that agency requires sentience, and therefore with *Exclusive Sentience Criterion*.

Still, sentientism about moral status has been challenged. Shelly Kagan claimed that "a possibility worth entertaining" is that:

> there are two potentially distinct bases for moral standing, sentience and agency, and either one is sufficient for grounding some sort of moral standing.[28]

This entails *Sentience Criterion*, but contradicts *Exclusive Sentience Criterion* and *Distinctive Sentience Criterion*.[29]

The capacity to have *preferences* or *desires*—a key component of agency—is sometimes viewed as a source of moral status.[30] This follows from accepting desire-satisfaction views of value, as well as the thesis that welfare subjectivity is sufficient for moral status. It does not require rejecting even *Exclusive Sentience Criterion*, since one could contend that genuine desires require phenomenal experiences. Other than those mentioned above, rejections of *Distinctive Sentience Criterion* are rare.[31] I did not find rejections of *Sentience Criterion*.

To summarize, basically everyone is a weak sentientist. Many also accept distinctive sentientism. Exclusive sentientism is debated, but popular. In what follows, "sentientism," left unqualified, typically refers to distinctive sentientism—a widely (but not universally) accepted view.

[26] Chalmers (2022); Roelofs (2023).
[27] Sebo (2017); Wilcox (2020).
[28] Kagan (2019, 19).
[29] See also Delon (manuscript).
[30] Shevlin (2020, 192–94).
[31] And usually concern illusionists or illusionist-adjacent thinkers (Dennett 1991; Carruthers 2019; Kammerer (2022a).

One problem of sentientism is that it is hard to apply to difficult cases. Some think that phenomenality plays a distinctive ethical role, but still fear that this does not help answer hard questions about the distribution of well-being and moral status, since the distribution of phenomenality itself is hard to know.[32] Marian Dawkins saw this problem as a reason to search for alternatives to sentientism.[33] She suggested treating factors like desire-satisfaction, rather than phenomenality, as the basis of well-being, as the resulting view is easier to apply to non-verbal animals: it is easier to know what they want rather than what they experience, or whether they have phenomenal experiences at all. Still, sentientism could be both true and hard to apply.

9.5. Assessing illusionism's ethical implications: Two key questions

Phenomenal realism and sentientism are intuitive and widely accepted: in our intuitive ethical picture, the phenomenal exists and plays at least a distinctive ethical role. If so, illusionism certainly has revisionary implications for the *ground* of well-being and moral status. Given illusionism, those cannot be grounded as they are grounded in our intuitive picture.

Does illusionism also imply revisions for the *distribution* of well-being and moral status—regarding what has which well-being and moral status? This depends on our answers to two key questions. First, which link between phenomenality and value and moral status do we admit? Second, what, aside from phenomenality, grounds value and moral status?

Consider the first question. Illusionists conceding exclusive sentientism are led to *nihilism* about well-being and moral status: nothing is good or bad for anyone, and no one deserves moral consideration. This is extremely revisionary. Those who already find arguments for ethical nihilism convincing might find this acceptable.[34] They could also think that ethical nihilism aligns well with illusionism: for instance, both views can be supported with debunking arguments. Still, notice that nihilism about well-being, at least, is often judged even more implausible than nihilism about morality.[35] Here,

[32] Sebo (2018); Shevlin (2020).
[33] Dawkins (2021, 98).
[34] For instance, the arguments supporting the moral error theory (Mackie 1977).
[35] Fletcher (2017).

I assume that ethical nihilism is a drastically revisionary position. A view implying it arguably contradicts manifest ethical truths and becomes vulnerable to a direct refutation.

Revisions are less extreme, but still significant, for illusionists conceding only distinctive sentientism. Focus on well-being. Illusionists accepting *Distinctive Phenomenal Value* must admit that the prudential value of many of our mental states is quite different from what we intuitively think. We think it is very good for us to receive pleasant massages, but it is not as good as we think: it does not involve phenomenal pleasure, which would have grounded distinctive value. We think it is very bad for animals to live in factory farms, but since their lives involve no phenomenal suffering, grounding a distinctive sort of disvalue, it is not as bad as we think. These are significant distributional revisions.

Which exact revisions must illusionists conceding distinctive sentientism accept? This depends on their answer to the second key question: What, aside from phenomenality, grounds value and moral status? Depending on their answer, they might or might not conclude, e.g., that a massage is still good for us and that the life of animals in factory farms is still bad for them, in virtue of non-phenomenal value-grounding features. Still, since they accept *Distinctive Phenomenal Value*, they must concede that the value involved is not the distinctive one that phenomenal states would have grounded. Drastic distributional revisions seem hard to avoid, except maybe if we suppose a coordinated downgrading of all value, where everything has its value proportionally diminished as we realize that phenomenality is non-existent, in a way that leaves all *comparative* prudential claims untouched. The same applies to moral status for illusionists granting *Distinctive Sentience Criterion*.

Those, like Strawson and Balog, who fear that illusionism's adoption would be a disaster and lead to disregarding human well-being and moral status probably do so because they endorse distinctive or exclusive sentientism, and assume that illusionists do too. As I will show, their fear is unwarranted, because illusionism undermines our reasons to accept distinctive or exclusive sentientism.

As to illusionists merely conceding *weak* sentientism, they might avoid distributional revisions: for them, non-phenomenal features could ground the value and moral status we thought was grounded in the phenomenal. Factory farming could be as bad for animals as we thought, not because it involves certain phenomenal experiences, but because it has non-phenomenal components grounding the negative value we thought came

from phenomenal suffering. Illusionists conceding *Distinctive Phenomenal Value* cannot say this: they must concede that some distinctive value admitted by our intuitive picture is absent from reality.

Still, illusionists conceding weak sentientism and denying distributional revisions must claim that, for instance, the negative value coming from living in a factory farm, which would be grounded in phenomenal pains if animals had them, is really grounded in non-phenomenal features—say, functional features of the animals' mental lives. However, these other features would also exist if animals had phenomenal pains. This suggests that, if farmed animals had phenomenal pains, there would be *two* grounds for the same amount of value: one phenomenal, one non-phenomenal. This generalizes to other cases. This apparently entails a systematic constitutive overdetermination of value and moral status, which seems like an amazing coincidence. So, illusionists conceding weak sentientism are not committed to distributional revisions, but they might still struggle to avoid them. Ultimately, again, which revisions they admit depends on their answer to the second key question: What, aside from phenomenality, grounds value and moral status?

The same applies even to illusionists rejecting weak sentientism. They avoid the amazing coincidence problem, but they are not guaranteed against distributional revisions. Whether they are led to these revisions or not depends on their answer to this second key question.

9.6. Illusionism and sentientism

In this section, I tackle the first of the two key questions: Which link between phenomenality, on the one hand, and value and moral status, on the other hand, should we admit? Which sort of sentientism, if any, should we concede? I argue that illusionists should concede weak sentientism, but maybe not distinctive or exclusive sentientism.

A. Conceding weak sentientism

Come back to the *Pure Suffering* thought experiment. When Piotr phenomenally experiences pain, it is bad for him: *Phenomenal Value* is extraordinarily intuitive. Does illusionism suppress or undermine this intuitive support?

First, how is this intuition generated? Earlier, in Chapter 6, I hypothesized that phenomenal concepts of valenced experiences characterize them as having something like primitive value. If so, our intuition that Piotr's phenomenal pain is bad is *conceptual*: it stems from the way our concept of phenomenal pain characterizes its referent. In fact, I think that, plausibly, *Phenomenal Value* is an implicit conceptual truth—a truth which, given illusionism, concerns something non-existent, and remains innocuous.

The religious analogy helps. Imagine Friedrich, a convinced atheist. He concedes that *eternal life in Heaven* would be good for us. Our well-being would be very high in a state of eternal bliss after bodily death. Friedrich could even concede this as a conceptual truth if he admits conceptual links between bliss and value. Still, he would find this conceptual truth irrelevant to deciding what is good for us in reality, as it concerns something unreal and physically impossible. The heavenly afterlife is good, but this does not matter for us who live in a Heaven-less reality. Similarly, illusionists should concede *Phenomenal Value*, but treat it as essentially irrelevant.

Some illusionists could deny *Phenomenal Value*. Here are two strategies they could use. First, they could claim that our intuition of *Phenomenal Value* disappears on reflection. Second, they could claim that, even if it remains, illusionism undermines its evidential value: given illusionism, we should not trust this intuition.

Start with the first strategy. Illusionists could concede that we have a robust intuition that *Pure Suffering* is bad, but that this does not really support *Phenomenal Value*. Drawing inspiration from previously mentioned attitudinal approaches to valenced states or desire-satisfaction theories of well-being, they could contend that, intuitively, what is bad about Piotr's pain is not his phenomenal feelings, but that he *dislikes* his feelings or that he has a frustrated desire for them to stop. If so, *Pure Suffering* is bad, but this fact only seems to support *Phenomenal Value* because we confuse pain's phenomenal character with what really makes pain bad: the fact that we dislike it and have a frustrated desire that it stops. To make this plausible, one can notice that, intuitively, if Piotr were a masochist, if he *liked* his pain and desired it, the pain would not really be bad for him. This suggests that what is bad in his pain is the dislike or the frustrated desire, not the phenomenal feel.

I find the reasoning unconvincing. Concede that Piotr's pain is intuitively bad because of his dislike or frustrated desire. However, this seems most plausible in a sense of "dislike" or "desire" expressing *phenomenal* concepts. Consider *Pure Zombie Suffering*:

Pure Zombie Suffering: Zombie Piotr is physically and functionally identical to Piotr, but without any phenomenal experiences.

Our intuition that Zombie Piotr's pain is bad is not as overwhelming as with Piotr's pain. Some might even intuit that Zombie Piotr's pain is not bad at all for him. Our overwhelming intuition that Piotr's pain is bad might depend on our apprehension of phenomenal features—maybe, phenomenally conceived dislikes, or phenomenally conceived unsatisfied desires.

We can make the same point with hedonic inversions.[36] Take Natasha, who has a condition opposite to Piotr's: every night, she enters bliss for one hour (*Pure Bliss*). Then, take Andrei, who each night enters a state phenomenally like Natasha's *Pure Bliss*, but functionally and physically like Piotr's *Pure Suffering*. Intuitively, Andrei's *Inverted Bliss-Suffering* is not as bad as *Pure Suffering*. Therefore, if what is overwhelmingly intuitively bad about pain is that we dislike it or desire it to stop, it is in a sense of "dislike" or "desire" expressing phenomenal concepts.

Concede that *Phenomenal Value* is robustly intuitive. Does illusionism undermine the evidential value of this intuition?

One could argue that it does, e.g., because we should always distrust value-intuitions about the non-existent. However, this seems inelegant. This is not how we usually treat value-intuitions. Friedrich the atheist easily admits that eternal life in Heaven would be valuable, even if it does not and cannot exist. He seems right to do so.

Illusionists are better off simply conceding *Phenomenal Value*. For broadly similar reasons, they should concede *Sentience Criterion*. They should concede weak sentientism, as an innocuous conceptual truth bearing on something non-existent—the same way atheists easily concede that heavenly life would be good for us and that angels would have moral status.

B. Doubting distinctive and exclusive sentientism

In contrast, embracing illusionism gives us reasons to distrust our intuitions supporting distinctive sentientism and exclusive sentientism.

First, and even if this is not key to my argument, note that our intuitions supporting distinctive sentientism and a fortiori *exclusive* sentientism are

[36] Kahane (2009, 334); Kammerer (2020, 914).

not as strong as those supporting weak sentientism. For instance, the claim that it is morally acceptable to let a billion zombies die when we could save them at a small cost is, in my mind, counterintuitive. Yet, it is entailed by *Exclusive Sentience Criterion*.

Still, concede that distinctive and exclusive sentientism are intuitive. Illusionism undermines the evidential value of these intuitions. These intuitions are not conceptual, but depend on biases and implicit presuppositions whose falsity is implied by illusionism. This is what I will now argue, focusing on our intuition of *Distinctive Phenomenal Value*. I assume that the same applies to our intuition of *Distinctive Sentience Criterion* and, a fortiori, to our intuitions supporting exclusive sentientism.

Why think that our intuition of *Distinctive Phenomenal Value* is not conceptual? Mostly, because it can be explained well otherwise.

In Chapter 6, I explained our intuition of *Exclusive Phenomenal Value*, and a fortiori of *Distinctive Phenomenal Value*, by our antecedent acceptance of the *normative relocation*. Remember the normative relocation story. Affective cognition initially led us to believe that many things had primitive value. We struggled to locate this value in the non-phenomenal, but not in the phenomenal world—a realm of primitive qualities intrinsically presented to subjects, where things are as they seem. Grasping phenomenal experiences as introspectively obvious, we found it hard to envision that phenomenal experiences could introspectively seem valuable but fail to be such, while this was unproblematic for non-phenomenal things. We came to view the phenomenal as *the* locus of intrinsic value, the non-phenomenal having at most instrumental value. Now, as we usually tend to accept this normative relocation, we think of *Zombie World* as valueless, and a fortiori, we think of *Zombification* as featuring a dramatic loss of value.

Given illusionism, the normative relocation was misguided: the phenomenal is not real. We cannot solve the value-location problem by placing it in the phenomenal realm. Note that illusionism does not dictate a view of value: it is compatible with relocating it in the mind, just not in the phenomenal mind.

For illusionists, when we view *Zombie World* as a value desert, we are like a dupe fooled by the sleight of hand by which a skillful conjurer plays a Three-card Trick. The conjurer makes us believe that a card turned upside-down, on which our attention focuses, is *the valuable one*: this is the normative relocation step. The card is set aside: this is when we think of zombies, or of illusionism being true. We then become wrongly convinced that the remaining

cards are unimportant: this is when zombie worlds or illusionist worlds seem valueless.

Admitting our mistake, we can look at worlds devoid of phenomenal experiences with a fresh gaze. They do not result from the subtraction of the locus of value. They are worlds with loving, hating, struggling, striving, generous, sadistic, wise, and foolish creatures, whose lives can go very well or very badly.

Again, the religious analogy helps. When we view *Zombie World* as valueless, we are like the religious thinkers denounced by Nietzsche, who *devalue* our world—our ordinary natural world—[37] in the name of an imagined superior, hidden, and purely intelligible "true" world—an "afterworld" (*Hinterwelt*).[38] Learning that there is no hidden intelligible world, they view the remaining ordinary world as valueless, but they only do so because they view it as deprived of the value they first placed in the hidden intelligible world. As Nietzsche puts it, the shadow of God lies over us "thousands of years" after God's death, and "we still have to vanquish his shadow, too."[39]

If I am right, our intuition that the phenomenal plays an exclusive (and a fortiori a distinctive) role for value stems from an antecedent implicit normative relocation made illegitimate by illusionism. Accepting illusionism, we should backtrack on the normative relocation and cease to have these intuitions. Given this diagnosis, these intuitions are not compelling nor robust, and should disappear on reflection.

However, things are trickier, as phenomenal realism itself is intuitively compelling. This makes it hard to fully accept illusionism at the intuitive level and to robustly backtrack on our antecedent implicit acceptance of the normative relocation, in order to generate new intuitions. Therefore, while our intuition that the phenomenal plays an exclusive role for value is cognitively penetrable, it might still behave mostly as if it were not, as it stems from beliefs that are themselves robustly intuitive. However, as these beliefs are (or depend on) false realist presuppositions, the resulting intuition is not to be trusted. Embracing illusionism undermines its status. The same applies to the moral status intuition.

Our intuition of *Distinctive Phenomenal Value* has another source. We tend to accept *Phenomenal Value* (plausibly, a conceptual truth) and phenomenal realism (a false, but compelling view). Therefore, we think of zombies

[37] Nietzsche ([1888] 1954b).
[38] Nietzsche ([1883] 1954a).
[39] Nietzsche ([1882] 1974, §108).

as *different* from us and lacking something valuable that we have. Imagining the sisters of *Zombification* before zombification, we imagine beings like us. Imagining them zombified, we envision them as deprived of something valuable that we possess. We view their lives as less valuable than ours.[40]

This does not yet explain why we view them as *distinctively* less valuable. However, a near-universal and unavoidable bias then kicks in: status quo bias. As we think we are sentient, we view the value of sentient beings' lives as our baseline. Anything that deviates negatively from this baseline seems *considerably* less valuable. Similarly, people with certain goods underestimate the value of lives without those goods. Think of the cliché of the rich Westerner who wrongly feels that lives without modern comforts, medicine, vacations, etc., are barely worth living, while those who live such lives often disagree. Think of the "disability paradox"—non-disabled people overestimating the welfare loss disability brings.[41]

This generates the intuition that zombies' lives are considerably less valuable than ours. Again, as this intuition is generated by a false belief in phenomenal realism, as well as status quo bias, it is untrustworthy. Given illusionism, the sisters of *Zombification*, pre-zombification, are *different from us.* They have something extra, which does not exist in reality. Zombification renders them like us. Does their well-being change? Maybe. Maybe their lives were better or worse when they had this mysterious extra feature, but there is little reason to believe that the difference was really so significant, and what matters anyway is that their 'ordinary' post-zombification lives have all the pain and pleasure, joy and sorrow, love and hate humans have always had. These are full lives, not botched ones.

Again, a religious analogy is helpful. Fyodor thinks that we have immortal souls directly created by God. He feels that purely physical soulless 'humans' would have impoverished lives, with little to no value, akin to mere machines. His friend Friedrich the atheist thinks that we *are* and always were soulless. He does not see the lives of soulless creatures as worthless; it is his baseline. Having immortal souls created by God might have been good for us—who knows? But these do not exist and could not (physically) have existed: they cannot shed a negative light on our soulless lives. Friedrich does not miss what he knows he never had.

[40] Technically, this conclusion is avoidable, and we could think that zombies can have lives as valuable as ours if we ourselves always instantiated two sufficient grounds for the same value (one phenomenal, one non-phenomenal). However, as noted above, this would be an amazing coincidence.

[41] Albrecht and Devlieger (1999).

Given this diagnosis, the intuition of *Distinctive Phenomenal Value* is cognitively penetrable. If we were to give up phenomenal realism, it would disappear, and we would be like Friedrich. However, again, phenomenal realism is compelling. Even illusionists relapse into implicitly accepting it. They thus feel that zombies lack something valuable we have. As status quo bias kicks in, they view zombie lives as having little to no value. As they remember that they themselves are not different from zombies, they can be horrified at the idea that their lives have so little value. However, they only reached this point when stuck in a biased disvaluing loop whose evidential value is undermined once they understand it.

Illusionists will typically fail to be like Friedrich, and react like Fyodor, the believer, who became an atheist after years of faith but suffers from God-nostalgia, as he keeps strong religious intuitions that make him see his own soulless live as relatively valueless. Our realist intuitions have the same effect, but as we understand this effect, we can distrust our intuition that the phenomenal plays a distinctive or exclusive role for value. The same applies to moral status intuitions.

To summarize, the intuitions supporting distinctive and exclusive sentientism are not conceptual. They have two sources. One is the implicit antecedent operation of the normative relocation, requiring realist presuppositions. The second is status quo bias combined with realist presuppositions and conceptual intuitions of *Phenomenal Value* and *Sentience Criterion*. Even convinced illusionists might persistently feel the intuitive pull of distinctive and exclusive sentientism, but they will have good reasons to distrust these intuitions. So, at this stage, illusionists should concede weak sentientism, but they do not need to concede distinctive or exclusive sentientism. One path leading them to drastic ethical revisions is avoided.

C. Toward alternatives to sentientism

What precedes does not yet establish that illusionists *should* ultimately reject distinctive or exclusive sentientism, nor that they escape drastic revisions. Whether this happens depends on their view of what grounds value and moral status outside of the phenomenal. Those who think that *nothing* non-phenomenal grounds value and moral status are led to exclusive sentientism and ethical nihilism—a drastically revisionary view. Still, illusionists are likely

to adopt ethical non-sentientism, on which something non-phenomenal grounds value and moral status—at least, as long as there is at least some justification for at least some version of ethical non-sentientism, as I will show is the case. Adopting ethical non-sentientism is compatible with accepting weak or distinctive (but not exclusive) sentientism as I defined them. Which non-sentientist ethical view is chosen determines the revisions one accepts. Illusionism does not dictate any given non-sentientist ethical view. In the following sections, I examine a few tempting options open to the illusionist.

Before turning to this, let us examine ethical objections to illusionism. Given what we have established, what can be said about them?

Illusionism alone entails no distributional revisions. It certainly does not entail ethical nihilism. It does not imply that pain is not bad for us, nor that humans have no moral status. Alone, it does not contradict putative ethical data. It might do so when conjoined with distinctive or exclusive sentientism, but those seem deniable. From putative ethical data about manifest ethical truths, we cannot deduce the falsity of illusionism alone. Of course, illusionism, combined with some specific ethical views, could entail drastic revisions. Assuming ethical data, we could rule out the conjunction of illusionism and these views, but illusionists could always respond by adopting alternative ethical views. Therefore, the ethical direct refutation of illusionism only succeeds if *all combinations* of illusionism with non-sentientist ethical views contradict ethical data, but this seems impossible, since illusionists could always adopt the ad hoc view that value and moral status are primitively distributed to fit whatever counts as ethical data. Of course, this sort of ad hoc view is implausible. Now, if all combinations of *plausible* ethical views with illusionism contradict ethical data, illusionism falls prey to an inference to the best explanation of ethical data. It struggles to explain ethical data, and only does so in conjunction with some implausible view. To know whether this more modest anti-illusionist ethical argument succeeds, we must assess non-sentientist ethical views.

9.7. Quasi-sentientism

I will examine four types of non-sentientist ethical views: quasi-sentientism, the valuable-illusion view, preference views, and objective views. These options do not exhaust possibilities, but hopefully they capture the most tempting ones.

Start with quasi-sentientism. For quasi-sentientism, quasi-phenomenality plays the role played by phenomenality in sentientism. Quasi-phenomenal experiences, notably valenced ones, play a distinctive (if not exclusive) role for value. Quasi-sentience plays a distinctive (if not exclusive) role for moral status.

Quasi-sentientism roughly maintains the intuitive distribution of value and moral status: the value and moral status we thought were grounded in phenomenality turn out to be grounded in quasi-phenomenality, but their distribution remains roughly the same: at least in the human case, quasi-phenomenality is present where we thought phenomenality was present.

The religious analogy helps. Consider again our deeply religious Fyodor. He thinks that something should be done if and only if God commands it, and in virtue of God's command—and he thinks he knows what God commands. He then stops believing in God and embraces the following view: something should be done if and only if it is *the type of thing that his God-command beliefs tracked*. Fyodor had a change of view about the grounds of duties, but not about their distribution. Illusionists endorsing quasi-sentientism take a similar stance.

Ethically speaking, quasi-sentientism neuters illusionism. Illusionist might be true, ethics can carry on as previously. This makes the view attractive. Versions of quasi-sentientism have been defended or favorably envisioned on several occasions.[42] Moreover, quasi-sentientism is substantively equivalent to a view *verbal* weak illusionists would call "sentientism." Since both verbal weak illusionism and the "sentientism" label are attractive to some degree, quasi-sentientism is substantively equivalent to a position likely to be popular beyond the ranks of strong illusionists.

Quasi-sentientism faces at least two significant problems: the determinacy problem and the justification problem. Start with determinacy.

In Chapter 8, I argued that the distribution of quasi-phenomenality (and maybe its basis and extension) is likely to be widely indeterminate. This raises an issue for quasi-sentientism.[43] Plausibly, value and moral status should be determinate. Some think that they should *always* be determinate: something is good for a creature or not; a creature deserves certain moral consideration or not.[44] There are reasons to doubt this strong thesis: a

[42] By illusionists, or illusionism-adjacent thinkers (Dung 2022; 2024; Frankish 2024; Muehlhauser 2017; Veit and Browning 2023).
[43] This is notably inspired by Carruthers (2019).
[44] Dworkin (2011).

dose of ethical indeterminacy about borderline cases seems hard to avoid.[45] Still, it remains very intuitive that value and moral status are not *widely* indeterminate. Views rendering value and moral status widely indeterminate are deeply uncomfortable.

If the distribution of quasi-phenomenality is probably widely indeterminate, *exclusive* quasi-sentientism implies, for instance, that there are probably many animals or possible AIs such that their being welfare subjects or having moral status is indeterminate. This is uncomfortable because ascriptions of well-being and moral status play a key role in moral decisions. If they are widely indeterminate, we face what Jonathan Birch called a "meltdown of practical deliberations," which we can only avoid by adopting unpleasantly arbitrary principles.[46]

Superficially, *distinctive* quasi-sentientism fares better. It admits that features other than phenomenality ground value and moral status, and we might hope that these features determine value and status when quasi-phenomenality is indeterminate.[47] However, to do the job, the features must be sufficiently widespread and themselves determinate. Worse: to plausibly eliminate indeterminacy of value and status when quasi-phenomenality is indeterminate, they must ground *comparable* sorts of value and status as quasi-phenomenality, which contradicts the distinctive quasi-sentientist idea that quasi-phenomenality is distinctively significant.

This determinacy problem is serious, but not fatal, since quasi-sentientists can still hope that quasi-phenomenality will turn out to be mostly determinate—something I argued is unlikely, not impossible.

Second, quasi-sentientism also encounters a *justification* problem. Given illusionism, we do not have good reasons to believe that quasi-phenomenality grounds distinctive value and moral status. If we think we do, it is probably because we do not fully realize that quasi-phenomenality *is not* phenomenality.

Let me explain. Consider first the reasons realists have to think that phenomenality plays an ethical role. Earlier, I claimed *Phenomenal Value* and *Sentience Criterion* are conceptual truths: reflecting, for instance, on phenomenal pain, we see a priori that it is bad for the subject. However,

[45] Shafer-Landau (1994) suggests that accepting forms of moral indeterminacy helps the moral realist account for moral disagreement. Dung (2024, 24) cites Constantinescu (2014, 159–60), who deploys a sorite argument for indeterminacy in the moral domain.

[46] Birch (2022a, 809–13).

[47] Dung (2024, 16).

reflecting on *quasi-phenomenal pain* does not yield a similar outcome. Our concept of quasi-phenomenal pain is the concept of a state mischaracterized as phenomenal pain. It is not a priori that this state is bad for the subject. The religious analogy helps again. Our concept THINGS GOD COMMANDS might plausibly characterize its referent as dutiful. Our concept THINGS MISCHARACTERIZED AS COMMANDED-BY-GOD certainly does not.

So, contrary to what happens with phenomenality, the putative ethical role of quasi-phenomenality is not justified by conceptual reflection. Is it justified otherwise? *Introspection* is a candidate. Does introspection give us reasons to believe quasi-phenomenality plays an ethical role? Luke Muehlhauser suggested so:

> [My] intuitions about (e.g.) the badness of conscious pain and the goodness of conscious pleasure were never dependent on the "reality" of specific features of consciousness that the illusionist thinks are illusory. Rather, my moral intuitions work more like the example I gave earlier: I sprain my ankle while playing soccer, don't notice it for 5 seconds, and then feel a "rush of pain" suddenly "flood" my conscious experience, and I think "Gosh, well, *whatever this is*, I sure hope nothing like it happens to fish!" And then I reflect on what was happening *prior* to my conscious experience of the pain, and I think "But if *that* is all that happens when a fish is physically injured, then I'm not sure I care."[48]

On this view, endorsed by Leonard Dung,[49] introspection justifies value-beliefs about mental states which could turn out to be phenomenal or quasi-phenomenal. This justification is not defeated when we learn that these states are quasi-phenomenal. If so, quasi-sentientism is justified.

Is this view of the introspective justification of value-beliefs plausible? Consider one of its implications. Given this view, when we are in pain, we are introspectively justified in believing *Badness*:

> *Badness*: "*This* is bad, even if *this* is not *thus*."

"*Thus*" expresses the phenomenal concept of pain. "*This*" expresses a neutral demonstrative concept referring to quasi-phenomenal pain in our

[48] Muehlhauser (2017).
[49] Dung (2022; 2024).

illusionist world—which would have referred to phenomenal pain had realism been true.

Suppose introspection justifies us in believing *Badness*. First, note that this justification cannot be uniquely secure. *Phenomenal experiences* might be introspectively obvious, but quasi-phenomenal states are not—quite the contrary, since they are different from what they introspectively seem to be. Our putative introspective justification for *Badness* does not have the safety we intuitively ascribe to our introspective justifications regarding the phenomenal.

This matters, since one thing making sentientism attractive is our impression that the ethical role of phenomenality is securely justified. This made the normative relocation (and thus exclusive and distinctive sentientism) tempting. The value of the phenomenal seemed certain, while the value of the non-phenomenal processes seemed always debatable. If introspection justifies us in believing *Badness*, it is not in the secure way that made sentientism so attractive. Instead, it is in a fallible manner, epistemically on a par, say, with the justification we thought affective cognition provided regarding non-phenomenal value before we endorsed the normative relocation.

Things get worse. Concede that introspection provides some prima facie, defeasible justification for *Badness*. Illusionism undermines it, as it states that quasi-phenomenal states are *systematically introspectively mischaracterized*—as phenomenal. This undermines the evidential status of introspection when it depicts *other* properties of our quasi-phenomenal states, like value-grounding properties.[50]

Against this, Dung noted that, while phenomenal introspection mischaracterizes quasi-phenomenal states as phenomenal, it is reliable in other respects: it reliably detects which states we are in.[51] We seem able to reliably introspectively discriminate between pain and pleasure, experiences of red and green, etc.

To know whether the evidential status of *ethical* introspection—the introspection of value-grounding or status-grounding properties—is undermined, we must know whether ethical introspection is more similar to *mischaracterizing* introspection (introspection as phenomenal) or to *reliable* introspection (introspective discriminations of quasi-phenomenal states).

[50] Kammerer (2022a).
[51] Dung (2024).

For Dung, ethical introspection is like reliable introspection and unlike mischaracterizing introspection, because it is metaphysically neutral and does not contradict physicalism.

However, *pace* Dung, ethical introspection resembles mischaracterizing introspection in striking ways, distinguishing it markedly from reliable introspection—at least, to the extent that it justifies *distinctive* quasi-sentientism.[52] To support distinctive quasi-sentientism, ethical introspection must depict quasi-phenomenality as grounding *distinctive* value or moral status—value or status considerably more significant than those grounded by other states and capacities. This implies that ethical introspection and mischaracterizing introspection (as phenomenal) are similar: both depict their targets as having non-ordinary, special properties that make their targets unique. By contrast, reliable introspection merely differentiates states from one another, without ascribing to them non-ordinary, special properties. Illusionism undermines the evidential status of ethical introspection.

Of course, other classifications put ethical introspection in the *reliable* category. Determining the right classification is hard. This is reminiscent of the *generality problem* for epistemological reliabilism.[53] Still, *ascribing properties that make the referent special* seems relevant in an epistemological context. I doubt that more relevant categories ultimately justify a counter-induction.

The previous dialectic seems complex, but the underlying idea is simple. For illusionists, quasi-phenomenal states are ordinary mental states introspectively mischaracterized as having special features—as phenomenal. This undermines other putative introspective reasons to believe that quasi-phenomenal consciousness has other special features, like special ethical features.

Quasi-sentientism is not justified by conceptual reflection or introspection. A last candidate justifier is empathy. Seeing my friend Paul screaming while holding his foot, my empathetic reaction might give me reasons to think that what happens to him is bad for him. Can this justify quasi-sentientism? Probably not. Empathy's verdicts do not seem precise enough to support specifically quasi-sentientism. They do not really tell me what is bad about Paul's state. Contrary to introspection, it is doubtful that empathy specifically tracks quasi-phenomenality. Besides, empathy is easily manipulable and sensitive to irrelevant factors: e.g., we have more empathy for those

[52] Kammerer (2024).
[53] Goldman and Beddor (2021).

who resemble us, or those we find attractive.[54] Even if empathy could justify specifically quasi-sentientism, the justification would be fairly weak.

9.8. The valuable-illusion view

The valuable-illusion view states that the illusion of phenomenality plays the role played by phenomenality in sentientism. Illusions of phenomenal experiences (notably of valenced ones) play a distinctive (if not exclusive) role for value. The capacity to have illusions of valenced phenomenal experiences plays a distinctive (if not exclusive) role for moral status.

The valuable-illusion view remains close in spirit to the intuitive sentientist picture. It partly conserves the intuitive distribution of value and moral status. We thought phenomenality was real and grounded value and moral status. It does not exist, but in many (though perhaps not all) places where we thought it did, there is an illusion of phenomenality, which made us believe in it in the first place, and this illusion grounds value and moral status.

To use our religious analogy: Fyodor the believer thinks that humans' well-being distinctively depends on their relation to God. When he embraces atheism, he concludes that their well-being in fact distinctively depends on their *illusion* of a relation to God—their false impression of being loved by God. Praying or going to mass is still good for us, not because it sustains our relation to God, but because it generates an illusion thereof.

The valuable-illusion view is often voiced in conversation, but I have never seen it defended in print. It is superficially attractive for two reasons.

First, like quasi-sentientism, though not as much, it limits distributional revisions. Still, this depends on how we view the illusion of phenomenality. If it is a permanent illusory self-monitoring process widely shared across species, its distribution might be comparable to the most intuitive distribution of phenomenality for realists, and the valuable-illusion view leads to few revisions. If it results from effortful processes requiring rare, sophisticated, and idiosyncratic capacities, it is scarce—plausibly, only humans have it. This might lead to serious revisions.

Second, the view stems from an intellectual move that seems legitimate. Learning that a situation involves an illusion often does not change our assessment of its value. Take *phantom limb pain*—subjects who feel pain in a

[54] Bloom (2018).

missing limb. It involves an illusion, but this does not influence how we view its value. A phantom-limb patient being told that her condition is not bad for her, as she only has an *illusion* of a damaged limb, would retort that the illusion is bad all the same. When told that phenomenal pain itself is illusory, a similar move ("the illusion is bad all the same") leads to the valuable-illusion view.

The valuable-illusion view encounters three problems. Two, like previously, concern justification and determinacy. The third concerns its overall plausibility.

First, justification. The attractiveness of the option rests on a confusion. The move supporting the view only seems legitimate because we usually apprehend illusions as grounded in phenomenal experiences, and because we are intuitive sentientists who think that the phenomenal is the locus of value. Phantom limb pain involves an illusion, but we think that it still involves the same phenomenal property as 'normal' pain, and that is why we conclude that it is as bad.

However, the same move is not as straightforward if we abandon realist and sentientist presuppositions. Without this move, the valuable-illusion view is hard to justify, roughly for the same reasons as quasi-sentientism. *Conceptual reflection* does not justify it: the concept of a non-phenomenal illusion of a valuable phenomenal experience does not characterize its referent as valuable. *Introspection* does not fare better: even conceding, for the sake of the argument, that we introspect illusory representations of phenomenal experiences as valuable, the justification provided by such *second-order* ethical introspection is undermined for roughly the same reasons mentioned earlier about first-order ethical introspection.

Second, indeterminacy. Earlier, I stressed that the *potential* distribution of the illusion of phenomenality is probably indeterminate: the valuable-illusion view, like quasi-sentientism, faces an indeterminacy problem. This one might be less pressing, since this indeterminacy becomes most plausible only for merely potential beings.

Finally, tying distinctive value and moral status to *illusory* representational processes seems simply implausible. Consider this thought experiment:

> *Disillusioning Surgery:* Paulus Kirchland is a convinced illusionist of the 22nd century.[55] Like other humans, he is under the illusion of

[55] Paul Churchland thought introspection is so flexible that, through learning and training, we could come to introspect our mental states as neural states (Churchland 1985).

phenomenality. One day, he signs up for a subtle form of brain surgery, which *frees him* from his mistaken tendency to introspect phenomenal states, and leaves all his introspective discriminative abilities intact. Post-surgery, Paulus is freed from the illusion: he discriminates his quasi-phenomenal states as well as before, but now correctly characterizes them, in introspection, as neuro-functional states rather than phenomenal states.

On the valuable-illusion view, Paulus's capacity for well-being and moral status greatly diminishes after surgery. This seems highly implausible. The surgery did not change him much, and to the extent that it did, it generated epistemic progress. More generally, it does not seem plausible to make ethical features depend so deeply on some of our representations being so specifically epistemically defective.

9.9. Preference views

I now turn to two other, more radical alternatives to sentientism: preference views and objective views.

Start with preference views. They state that what matters distinctively (if not exclusively) for value is the *satisfaction* of certain *preferences*. What matters distinctively (if not exclusively) for moral status is the capacity to have certain preferences.

Here, I use "preference" widely, to encompass desires, wishes, wants, the having of goals or aims, and comparable mental states. "Satisfaction" remains neutral between *objective satisfaction*—when a subject's preference happens to be satisfied—or *subjective satisfaction*—when a subject *takes* the preference to be satisfied, or has an attitude of *satisfaction*, *contentment*, or *liking* toward their situation.

Different views specify these concepts differently: preference views form a large family. One crucial point is that, since we seek alternatives to sentientism, preference-satisfaction must be understood as not essentially involving anything phenomenal. Some preference views are versions of quasi-sentientism, or the valuable-illusion view—e.g., if preference-satisfaction is understood as always involving something quasi-phenomenal, or an illusion of phenomenal experience.

Preference views seem intuitive about well-being. Almost as a matter of definition, what we want matters to us. Getting what we want seems good

for us. Preference views correspond to the desire-satisfaction conception of well-being described earlier. A combination of hedonism about well-being and attitudinalism about pleasure and pain,[56] on which pain and pleasure consist in having the right attitudes of liking or disliking, might also lead to a preference view.[57] Preference views are influential about animal welfare.[58]

That having preferences matter for moral status is also intuitive. Suppose you are kidnapped by aliens who want to keep you in a zoo. You try to convince them to release you.[59] You might first mention your sentience, but if this does not work, you might stress that you *care* about things, and that being kept in a zoo prevents you from obtaining those things—spending time with your loved ones, writing philosophy, reading poetry, traveling the world, etc. You would say so because, plausibly, caring about things makes you deserving of moral consideration.

As a conception of moral status, preference views are close to the influential idea that *agency* is sufficient for moral status.[60] As I understand it, having preferences is central to agency, even if it does not exhaust it: agency also requires at least the capacity to act on these preferences.

Preference views might form the most popular non-sentientist ethical option. I already mentioned Marian Dawkins. Guided by applicability concerns, she set sentientism aside and turned to a view of animal welfare incorporating key aspects of preference views.[61] Preference views are often chosen by illusionists or illusionism-adjacent thinkers.[62] Peter Carruthers defended something like a preference view of animal welfare against sentientism. Daniel Dennett presented a preference view as an upshot of his illusionism.[63]

Do preference views imply distributional revisions for illusionists? Most things that matter in our intuitive picture also matter given many preference views. Valenced experiences, such as pain and pleasure, have significant value in our intuitive picture, but also in many preference views, since they typically involve satisfaction or dissatisfaction of preferences. Massive *subtracting* revisions are avoided.

[56] Feldman (2002).
[57] For an argument that this version of hedonism is identical with a version of desire-satisfactionism, see the already mentioned Heathwood (2006).
[58] Jensen and Sandøe (1997); Appleby and Sandøe (2002, 289–90).
[59] This is inspired by Gruen (2011, 26); cited in Kagan (2019, 16). See also Kagan (2019, 28).
[60] Kagan (2019); Delon (manuscript).
[61] Dawkins (2021, chap. 8).
[62] Kammerer (2020; 2022a).
[63] Dennett (1991, 450).

Still, *additive* revisions might occur, depending on how we define "preference" and "satisfaction." Construed widely, preferences are widespread.[64] Unicellular organisms might prefer some situations over others and strive to reach them.[65] A relaxed understanding of preferences leads to counterintuitive revisions: intuitively, it is clear that unicellular organisms are not welfare subjects, nor moral patients. These revisions are avoided or tempered if only a restricted set of preferences matters, or if only a restricted set of preference-like processes count as genuine preferences. They can also be tempered if all preferences matter, but some matter much more than others. This seems to have been Dennett's view:

> [The] greater the scope, the richer the detail, the more finely discriminative the desires, the worse it is when those desires are thwarted.[66]

This seems plausible when examining various cases. Consider, on one extreme, the 'preference' of *Bacillus subtilis* for pH-neutral environments. On the other extreme, consider humans. They *care* about things: they enter a "complex set of cognitive, affective and volitional dispositions and states,"[67] by which they want certain things they judge important for them, want to maintain their desire into existence and have it shape their behavior, and are persistently guided by all these desires. Humans are also capable of *love:* they can care about others and their flourishing, for their own sake.[68] Between these two extremes—the preferences of *Bacillus subtilis* and human love—there is, intuitively, a highly significant relevant ethical difference.[69] The second obviously matters much more for well-being than the first. Humans' capacity to love matters much more for moral status than *Bacillus subtilis*'s capacity to prefer pH-neutral environments.

Still, if we embrace a preference view and want to avoid distributional revisions entirely, we might have to say that what distinctively matters are preferences linked to states we thought were phenomenal: quasi-phenomenal states. The result is a variety of quasi-sentientism, or a closely related view. It faces the earlier concerns about indeterminacy and justification.

[64] Shevlin (2020).
[65] Lyon and Kuchling (2021).
[66] Dennett (1991, 450).
[67] Frankfurt (1982, 262).
[68] Frankfurt (1998).
[69] Kammerer (2022a).

A better view might be one in which, first, certain preferences matter more than others (e.g., because they are more complex, persistent, coherent, ramified, etc.), and second, these preferences are often, though maybe not always, linked to quasi-phenomenal states. This tempers distributional revisions while potentially avoiding indeterminacy issues, since quasi-phenomenality could be indeterminate even if complexity, persistence, coherence, ramification, etc., are not. This also helps with justification issues if we assume that it is appropriately intuitive that preferences that are complex, persistent, coherent, ramified, etc., are distinctively significant.

Now, consider difficult cases, about which our intuitive picture hesitates: bees, lobsters, sophisticated AIs, etc. If we accept a preference view, we have to identify the relevant preferences in the concerned beings, to determine their moral status and assess their well-being. Precise verdicts about each case will depend on open empirical questions and on the exact preference view chosen, but general features of these verdicts are predictable.

First, preference views might often lead to ascriptions of welfare and moral status that are less anthropocentric than, say, quasi-sentientism. After all, our concept of the quasi-phenomenal refers to states mischaracterized as phenomenal *in us*: the concept is anchored in the human case. So, for a quasi-sentientist, architectural and material similarity to humans might end up ethically relevant. By contrast, it is arguable that our concept of preference is a folk-psychological concept mostly analyzable in functional terms. It might not be as straightforwardly anchored in the human case, and its functional analyzability will make it fairly natural (though not unavoidable) to treat, for instance, material similarity to humans as irrelevant when it comes to the possession of preferences.

Second, since plausible preference views will probably admit something like an ethically relevant graded hierarchy of preferences, preference views are likely to yield more gradual distributions of value and moral status than on the intuitive picture.

Preference views encounter difficulties. Some are the classical problems of desire-satisfaction views of welfare.[70] Sometimes, satisfying our desires does not seem good for us: think of a drug addict or a child raised in a cult and brainwashed to want to perpetuate it. One can respond that only the satisfaction of *comprehensive* and *informed* desires is good for us. Still, this view also faces counterexamples: think of an intelligent and well-informed

[70] This paragraph summarizes points taken from Crisp (2021).

mathematician who deeply wants to spend her life counting blades of grass.[71] Intuitively, having her desire satisfied might not be the best thing for her. Alternatively, think of someone who entered a *desire machine*, changing their desires so that they desire exactly what happens. The person in the machine has all their (comprehensive, informed) desires satisfied, but intuitively, their life does not go as well as it could.[72]

On top of these problems, one issue arises specifically for non-sentientist preference views—those that interest us here. It regards their *justification*. I claimed that, intuitively, preference-satisfaction matters, but this might be most true of *phenomenality-involving* preference-satisfaction. Intuitively, *Pure Zombie Suffering* is not as bad as *Pure Suffering*.

One can also distinguish between two senses of "desire" or "preference": one in which desiring something means being genuinely attracted to the thing, and one in which it just means being disposed to bring it about.[73] Genuine-attraction desires are, intuitively, the only ones whose satisfaction matters for well-being. However, one might add, these genuine-attraction desires intuitively require phenomenal experiences on the part of the desirer.[74] So, intuitively, creatures deprived of phenomenal experiences cannot have desires whose satisfaction matters for well-being. This threatens the justification of non-sentientist preference views.

In response, we might stress, like earlier when discussing distinctive sentientism, that our intuitions that zombies *lack* value and moral status are undermined given illusionism. Similarly, our intuitions that the desires had by non-phenomenally conscious beings are not those welfare-relevant genuine-attraction desires could be undermined. Still, again, where we end up when it comes to the *positive justification* of non-sentientist preference views is unclear. Empathetic intuitions possibly help: thinking of zombies with frustrated desires, unrequited loves, crushed hopes, we empathize with them,[75] see that their lives go badly, and understand that they deserve moral consideration. Still, empathy raises previously discussed difficulties. The resulting intuitive support might be much weaker than when we thought preference-satisfaction involved phenomenality.

[71] Rawls (1971, 432).
[72] Forrester (2024); for related considerations, see Dorsey (2021, 212).
[73] Parfit (2011); Heathwood (2019).
[74] Lin (2020).
[75] For Carruthers (1999), we can empathize with a creatures whose desires are frustrated even if the frustration is not phenomenally experienced.

Preference views also face a *determinacy* problem. When we intuit preferences as ethically relevant, we arguably do so with a pre-theoretical concept of preferences. We can suppose that it characterizes preferences in functional terms—e.g., in terms of states that interact in certain ways with belief-like states to cause behavioral tendencies. Still, this characterization is probably too rough to pick out a single specific property in all cases. Considering difficult cases, such as possible complex AIs highly dissimilar to humans, whether they have preferences might remain indeterminate. This brings us back to the determinacy problem faced by quasi-sentientism.

In the best-case scenario, a certain sort of answer to the *justification* problem also gives us an answer to the *determinacy* problem. Indeed, if intuitions supporting preference views are generated once and for all, with our rough pre-theoretical concepts, the determinacy problem arises—except in the lucky case in which our pre-theoretical concepts latch onto a single specific natural kind. However, if these intuitions can be generated, modified, corrected, and refined *along the way*, as we contemplate various cases which we describe with progressively more accurate specific conative concepts, the determinacy problem evaporates. Learning more about how difficult cases work, we can form refined intuitions about whether various beings' preference-like states matter. Sometimes, these states might resemble our preferences in some respects, but not others. For instance, there might be states that make the creature strive to attain certain conditions, without bringing anything remotely like our frustration when the striving remains unsuccessful. As we familiarize ourselves with these beings, we refine our intuitions about the ethical significance of their states, and our refined ethical intuitions can be sufficiently shaped theoretically so that the specificity of the preference-concepts used to articulate our ethical views keeps up with what is required to describe the cases without threatening degrees of indeterminacy.

This 'solution' is not obviously open to quasi-sentientists. They typically think that *introspection* is what justifies our beliefs in the ethical significance of quasi-phenomenality, but introspection only concerns our own states, not the states of others. We might extend its scope by *imagining* what it would be to introspect states we do not have—states that we could describe with the help of progressively more precise theoretical concepts. However, this creates a tension with what made introspection an attractive candidate to justify quasi-sentientism. The defense of quasi-sentientism I discussed earlier, due to Muehlhauser and endorsed by Dung, precisely supposes that

our changes of mind about what introspection targets leave ethical introspective verdicts unchanged. If so, we cannot make refined judgments of ethical significance by imagining introspecting states we cannot have, because ethical introspection—or simulated deployments of ethical introspection—is insensitive to theoretical descriptions of the introspected states. Finally, while endowing beings involved in difficult cases with introspection is a theoretical possibility, for instance, for sophisticated AIs, this 'solution' raises difficulties. It is hard to decide if and when we should trust the resulting introspective verdicts, since, in principle, it seems clear that value and moral status could be systematically misrepresented.[76]

9.10. Objective views

Objective views are another radical alternative to sentientism. On objective views, what matters distinctively (if not exclusively) for value is the attainment of certain objective goods. What matters distinctively (if not exclusively) for moral status are capacities to obtain these objective goods.

Something is an objective good when its obtention is good for the subject, independently of whether or not the subject desires it: objective views stand in contrast with preference views. Still, objective views can state that some objective goods involve elements of preference-satisfaction, as long as there is no requirement that the subject desires the objective good itself for it to be a good.[77] For instance, *being in a loving relationship* might be an objective good, good for individuals even if they do not want it as such. Still, this good requires *some* desires of the subject to be satisfied: arguably, I cannot be in a loving relationship without having *some* satisfied desires to interact with the person I love.

I focus here on "pure" objective views, contrasted with "pure" preference views, but hybrid views are available. For instance, what matters distinctively for value could be the satisfaction of certain preferences for things that have certain objective features.

Objective views seem intuitive about well-being. Asked about what makes a life *good* for someone, many would mention things like love,

[76] For an optimistic take on endowing current models with ethically relevant introspection, see Long (2023); Perez and Long (2023). Reasons for skepticism are presented in Kammerer and Frankish (2023a). Birch's "gaming problem" arises here (Birch 2024).

[77] Fletcher (2013, 217).

friendship, health, freedom, knowledge, and wisdom—objective goods. This corresponds to the influential "objective list" conception of well-being,[78] arguably the most popular view of welfare in contemporary philosophy.[79]

Some traditions promoted something like an objective view about moral status—e.g., the Kantian tradition, on which the capacity to engage in rational reflection—arguably, an objective good—grounds moral status. Generally speaking, objective views of ethics have been historically popular and can be associated with thinkers as diverse as Plato, Aristotle, Spinoza, and Nietzsche. Different objective views can specify different objective goods: it is a large family of views.

Since I explore alternatives to sentientism, the objective views relevant here cannot merely specify as objective goods things with an essential phenomenal component. Some objective goods must be obtainable without anyone involved having any phenomenal experiences. Objective goods might require, sometimes or always, quasi-phenomenal states, or illusions of phenomenal experiences. This would not create any difficulty: at most, the view would become a version of quasi-sentientism or the valuable-illusion view.

One problem is that, with our non-sentientist constraint in mind, objective views lose some of their intuitive character. Consider standard plausible objective goods: love, pleasure, freedom, or knowledge. We typically think that they involve phenomenal experiences, and this could be key to their being, intuitively, objective goods. Take zombie versions of love, friendship, freedom, or knowledge. Are these, intuitively, objective goods? Do they seem to make zombies' lives good for them? Maybe not as clearly. As earlier, when the same problem arose for non-sentientist preference views, the illusionist can stress that our intuitions that zombies lack well-being and moral status are undermined given illusionism. Again, it remains unclear whether we end up with sufficient positive justification for the thesis that zombie love, friendship, knowledge, etc., are genuine objective goods.

To avoid this problem regarding the *justification* of non-sentientist objective views, we can attempt to provide a more specific account of what makes certain things objective goods, hoping that this account will help justify the idea that various things devoid of phenomenal components are indeed

[78] Fletcher (2013); Rice (2013); Crisp (2021).

[79] Fifty-one percent of professional philosophers accept the view, or lean toward it (Chalmers and Bourget 2020).

objective goods. Here are two strategies to achieve this, among others: one Aristotelian and the other Platonistic.

The Aristotelian strategy is a perfectionist one. The idea is that something is an objective good for a creature when it consists in the exercise or development of the creature's natural or essential capacities. It might be intuitive that it is good for someone to exercise and develop their nature. If so, the view avoids the justification problem. If illusionism is true, the exercise or development of one's essential capacities does not involve anything phenomenal, and the perfectionist approach justifies a non-sentientist objective view.

Perfectionism is influential as a view of human well-being,[80] as well as animal well-being.[81] One disadvantage of perfectionism is its appeal to essences. Skeptics will object that they do not exist or lack ethical import.[82]

The Platonistic strategy also faces difficulties, but I find it more attractive. It tries to justify that certain things are objective goods for creatures, not by focusing on the creatures or their essences, but by looking at the *entity* or *feature* with which the creature enters into relation when it obtains putative objective goods.[83] Consider a candidate objective good like *knowledge*. When one has knowledge, one is in a certain relation to *truth*. The Platonist reflects on truth, grasps that truth is, in itself, good-to-know, and thus justifies the view that knowledge is an objective good. Similarly, reflecting on beauty, we can grasp that it is good-to-appreciate; reflecting on goodness, that it is good-to-love. We thus justify that the appreciation of beauty and the love of the good are objective goods.[84]

The Platonistic strategy avoids the appeal to essence. It plausibly respects the non-sentientist constraint: it concludes that what is good is to enter into certain relations with certain objective features of reality, while seemingly leaving the nature of these relations open and making it secondary compared to the objective features themselves. Still, many will find this sort of justification debatable. They will fear that it only leads to the glorification of idiosyncratic tastes.

[80] Bradford (2015); Crisp (2021); Wall (2021).
[81] Rollin (1993; 2011); Appleby and Sandøe (2002, 286–89).
[82] Dorsey (2010).
[83] The resulting view is sometimes also called "perfectionism," but it must be distinguished from the result of the previous Aristotelian strategy—e.g., as "objective goods perfectionism," opposed to the Aristotelian "human nature perfectionism" (Wall 2021). Here, I avoid calling Platonistic views "perfectionist."
[84] Prototypical examples of resulting views are, in my mind, Weil ([1947] 2003); Murdoch (1980).

Is an illusionist endorsing a non-sentientist objective view led to distributional revisions? This depends on the view at hand, but general difficulties are notable.

Many perfectionists struggle to accommodate basic ethical claims. For instance, pain is manifestly bad for us, but perfectionism does not easily account for this. Setting aside cases of nociceptive dysfunctions, bodily pain *prima facie* counts as a perfectionist *good*—an exercise of our essential nociceptive capacities and an appropriate deployment of affective cognition.[85] As the badness of pain is manifest, the resulting view might contradict our ethical data.

Some bite the bullet. Various ethical traditions—Nietzscheism,[86] some forms of Christianism[87]—do not see pain, in itself, as bad for us, but this seems hard to swallow. Overall, however, there are no reasons to think that all objective views imply that pain is not bad for us.

Many objective views imply that beings very different from humans could have welfare and moral status. Perfectionism suggests they do as long as they have an essence. Potential *additive* revisions are to be expected.[88] Some are implausible. If animals have an essence, why not also plants or unicellular organisms? What about artifacts—do they have an essence fixed by what their creator intended, and therefore a corresponding welfare? Perfectionists have to circumscribe ethically relevant essences. Platonistic views might also ascribe welfare and moral status to creatures very different from us, even some without valenced cognition, if they are capable, say, of knowledge.

Finally, many objective views plausibly suggest that more capable creatures, those that are capable of obtaining more objective goods than others, have correspondingly higher moral status, and it is not obvious that one can reap the intuitive consequences of this (humans having higher moral status than complex animals, themselves having higher status than simpler ones, etc.) while entirely avoiding unpalatable elitist ones (exceptional humans or super-intelligent AIs deserving more moral considerations than ordinary humans like us).

As previously with preference views, however, for an objective view to avoid distributional revisions entirely, it must say that what distinctively

[85] Perfectionist have answers to this problem, but whether they work is contentious (see Bradford 2021 for an extensive treatment; for a response to Bradford, see Fletcher 2022).

[86] See Delon (2024).

[87] "I should not love my suffering because it is useful. I should love it because it *is*" (Weil [1947] 2003, 80).

[88] Appleby and Sandøe (2002, 208).

matters are objective goods linked to states we thought were phenomenal—quasi-phenomenal states. It must become a version of quasi-sentientism, or something close enough, facing the concerns about justification—how can we justify that quasi-phenomenal states are so special?—and determinacy—how to deal with the likely indeterminacy of quasi-phenomenality? The most plausible objective views, in my mind, do not take this route, and admit a dose of revisions.

Objective views face problems. One, already discussed, concerns their justification given the non-sentientist constraint. Another is often voiced from the perspective of the proponent of preference views: objective views of well-being, which do not require welfare goods to be wanted by the subject, seem to alienate subjects from their own well-being.[89] This concern can motivate the adoption of a hybrid view, mixing elements of preference views and objective views.

One last inconvenient of objective views, which also concerns hybrid views, is that they do not seem well-positioned to handle difficult cases—bees, lobsters, robots, etc. The Aristotelian stance suggests determining the essence of the relevant beings to assess their welfare, but even if essences exist, there is no obvious way to determine the essence of a being, let alone its normatively significant essence. I find it hard to see how we could make normatively relevant justified verdicts about the essence of a bee, a robot, or an LLM. Difficult cases seem at least as hard as they were for sentientists.

The Platonistic stance suggests examining the beings' capacities to assess whether they can enter into the right relationship with the right objective features. Perhaps we can identify some sources of welfare and moral status in this manner, at least when we ourselves can establish the right relationship with the right objective features. We are capable of knowledge, and, arguably, we therefore understand that, for those capable of it, knowledge is good. To the extent that robots and bees can *know* things, this is good for them—of course, a lot depends on how we specify what counts as (normatively relevant) knowledge.

However, how are we to determine whether *other* relations to *other* objective features, of which we are incapable, are good? Think of how the question might arise with AIs. Imagine an AI with cognitive and conative capacities very different from ours, who can enter relation R with F—R and

[89] Something which has been noted regarding perfectionism in general (Sumner 1992, 10), or applied to animals (Appleby and Sandøe 2002, 287–88).

F being beyond our intuitive comprehension. How are we to judge—from the outside, so to speak—whether entering R with F is good for the AI? The Platonist's best hope is to assume that various goods must converge, or resemble each other. As a helpful analogy, imagine deaf parents whose hearing daughter is interested in music. Suppose they cannot hire a teacher. Still, witnessing their child playing with drums and flutes, they could, from the outside, estimate that what she does is good for her. They would notice that the child's activity has structural features signaling progress: the child does not merely repeat the same gestures, nor vary them randomly, but engages in progressive diversification accompanied by innovation and partial accumulation. They would also notice that the child's activity coheres with other activities that are good for her: as she practices music, she seems to be in a better overall psychological state than otherwise (more joyful, relaxed, etc.). Whether this idea of a convergence or resemblance of welfare goods solves all dilemmas for Platonistic views remains doubtful.

9.11. Non-sentientist ethics and illusionism

A. Assessing non-sentientist ethical views

I presented four versions of ethical non-sentientism. The valuable-illusion view is, in my mind, the least plausible. While I understand the attraction of quasi-sentientism, most of it disappears once we really accept illusionism and understand its consequences. The best non-sentientist ethical views are certain preference views, certain objective views, and certain hybrid views mixing elements of both.

My own preferred approach, for which I will not try to argue, is a hybrid view mixing elements of preference views and elements of a Platonistic objective view. It says that what matters for well-being are situations where one has certain things worth caring about (say, knowledge and love) *and* cares about them in ways deeply sensitive to what makes these things worth caring about—presupposing that one 'gets it' and grasps in the right manner (i.e., by appropriately caring about) what is good in, say, knowledge and love.[90] What

[90] This comes close to the sort of hybrid view of welfare which make well-being a matter of "enjoying the good," or "having satisfied desires for the good." See notably Woodard (2015, 164–65), mentioning Robert Adams, Joseph Raz, Richard Kraut, and Shelley Kagan as recent defenders of

matters for moral status is the capacity to have things worth caring about *and* care about these things in ways deeply sensitive to what makes them worth caring about.

Given this view, humans and animals have moral status because they have some of these capacities. Pain is bad for us because one of the best things for us is the maintaining of a loving relationship with others, ourselves, and the whole of reality, and that pain is one of the processes that are, constitutively and causally, most opposed to this relation. A given intense pain accompanied by rage, hate, and resentment is overall much worse for us than a pain of the same intensity, which we treat with calm acceptance, and during which we continue to love and forgive all the same. I think that, on reflection, this is fairly intuitive.

All non-sentientist ethical views, even those I judged plausible, encounter a challenge when it comes to their justification. Of course, some think that *every* theory of well-being,[91] or moral status, can be found lacking when it comes to ultimate justification. Still, I am inclined to concede that ethical sentientism prima facie seems uniquely well-justified. My view explains why: we start off with strong phenomenal realist tendencies and characterize our valenced experiences as instantiating something like primitive value. We also represent phenomenal experiences as introspectively infallibly self-revealed and as they introspectively seem to be. All this makes it extraordinarily difficult to intuitively envision that we do not have the valuable valenced experiences we seem to have. This is why we have such a strong intuition that we have ethically significant phenomenal experiences.

Once we understand that the phenomenal does not exist, we must probably also admit that no ethical views can have the sort of justification ethical sentientism seemed to have. In the first stage of grief, we might reject all ethical views, but in further stages, we appreciate non-sentientist alternatives and their—moderate, fallible, questionable—justification. The religious analogy helps us again. For someone raised religious, the first reaction to atheism might be to conclude that everything is permitted: after all, God gave us no command. Further reflection often leads to different conclusions. Most contemporary atheists are *not* like Ivan Karamazov, let alone Pavel

this sort of views. Woodard also stresses that, in the Appendix of *Reasons of* Persons, Derek Parfit considers a closely related view favorably (Parfit 1987, 501–2). See also the view that well-being consists in "fitting happiness" (Rossi and Tappolet 2022).

[91] Fletcher (2013, 217–18).

Smerdyakov. Figuring out the right non-sentientist alternative, and how to apply it to difficult cases, remains a formidable challenge. Given the current progress in the field of artificial intelligence, calling for systematic and justified ways to correctly ascribe welfare and moral status to beings very different from us, it is also an extremely pressing one.

B. Assessing ethical objections to illusionism

Earlier, I showed that illusionism in itself has no ethical consequences. The ethical direct refutation of illusionism fails. One can always devise an ad hoc ethical view which, combined with illusionism, yields the 'right' verdicts about value and moral status. Still, I conceded that the inference to the best explanation of ethical data would work if there were no plausible non-sentientist view of ethics which, combined with illusionism, avoids contradicting ethical data.

I have shown that various non-sentientist ethical views are available, and at least some of them are plausible. Some can avoid radical revisionary consequences and account for ethical data, e.g., regarding the badness of pain or the moral status of humans and animals. If so, the inference to the best explanation of ethical data against illusionism fails. Those who feared that the adoption of illusionism would lead to a prudential and moral disaster can be reassured. There can be ethical illusionists, just as there can be virtuous atheists.

Still, the non-sentientist ethical views I judged the most plausible—preference views, objective views, and hybrid views, such as my preferred view—probably lead to *some* distributional revisions regarding value and moral status. This means that illusionists are probably led to some ethical revisions. This is unsurprising. We might dislike ethical revisions, but it would be remarkable, if not suspicious, if illusionism, a highly revisionary conception of our mental life, did not suggest any ethical changes.[92] The religious analogy makes this vivid. Imagine someone who comes to believe that, contrary to what they thought, there is no personal God—no superior Being able to make commands regarding, say, diet or sexual behavior. This person would probably not renounce all moral beliefs or actions, but it would still be quite surprising if their religious change of mind had no impact on their

[92] Kammerer (2020, 919).

moral views. Virtuous atheists do not kill, but few of them have deep moral objections to work on Saturdays or pre-marital sex.

C. The ethical track record of non-sentientism

Before closing, I want to address two worries. The first concerns the ethical track record of non-sentientism.

Looking at my discussion of non-sentientist views, one could fear that we are trying, in an ad hoc fashion, to re-derive sentientist distributional claims in a non-sentientist framework. Arguably, what led us to 'discover' that some creatures have moral status and valuable lives was only the antecedent adoption of a sentientist perspective. Pushed further, a worry is that ethical sentientism has a better track record when it comes to establishing what we now admit as ethical truths. Non-sentientism only accommodates these admitted truths ex-post.

One thing making this idea of a better sentientist track record attractive is that, in the last decades (and maybe going back to Bentham), arguments for the ethical treatment of welfare subjects different from us, like non-human animals, were often put forward in sentientist terms.

Even granting that such track record arguments can succeed in the ethical domain,[93] it is not obvious that non-sentientist views only accommodate sentientist conclusions, nor that non-sentientist views have a poor ethical track record, even if we focus on the treatment of non-human creatures. To give a striking example, consider what the Arabic-speaking Andalusian philosopher Ibn Tufayl wrote in the 12th century, in his philosophical novel *Hayy Ibn Yaqzān*:

> There seemed to be three sorts of food: plants that had not yet reached peak maturity, [. . .] fruits of plants that had complete their life cycles [. . .] and animals, terrestrial and marine. [. . .] Feeding on them would unavoidably cut them off from their own fulfillment and prevent them all from achieving their intended purpose [. . .]. The answer, apparently, was, if possible, to give up eating completely. Unfortunately, [Hayy] could not do so because not eating tended to make his own body waste away, which was [worse], since he was superior to those other beings whose destruction meant his

[93] See Dung (2025) for a recent discussion.

> survival. So Hayy chose the lesser of evils [. . .]. Thus he could eat such things as the meat of fully ripened fruits. [. . . .] If it was hard to find fruit with nourishing meat [. . .] he would heave to eat [. . .] green vegetables [. . .]. If none of these were available, then he must eat meat or eggs.[94]

Ibn Tufayl had an (Aristotelian and theistic) objective view of ethics, on which what matters ethically is the fulfillment of beings' essential nature. Given this view, he ascribed well-being and moral status to various non-humans, at a time when this issue was not widely considered, which speaks in favor of the ethical track record of non-sentientism. Incidentally, Ibn Tufayl suggested what could count as 'additive' revisions to our current intuitive picture, as he claimed that plants deserve moral consideration: Ibn Tufayl's hero, Hayy, has a duty to eat fruits over green vegetables. This gives weight to my previous suggestions that objective views easily lead to additive revisions.

D. Ethics, mystery, and humility

A last worry regards our ethical attitude and our sense of mystery. We view phenomenality as mysterious. This creates problems for sentientism, by making it hard to apply. Still, it has a virtue. It furthers an attitude of fearful respect when considering ethical issues raised by various beings' mental lives. Sentientists think that phenomenality is what is ethically significant, but phenomenality is also something we apparently cannot easily explain or detect, and therefore something we cannot easily replicate, control, and modify at will.

It might be good to view what ethically matters as something that might escape our control and challenge our certainties. It is a (broadly correct) cliché that modernity issued a technicist era: we primarily apprehend the world as populated with measurable and controllable entities. Various concepts describe this, from Weber's disenchantment and iron cage to Heidegger's Enframing (*Gestell*). Treating phenomenality, viewed as mysterious, as the source of ethical significance, helps curb our arrogance when dealing with beings that matter, or could matter for all we know.

[94] Ibn Tufayl (2009, 110–12).

The case of AI consciousness makes this manifest. A sentientist, seeing phenomenality as mysterious, fears that engineers could unknowingly create conscious AIs with well-being and moral status, whom we might mistreat.[95] A sentientist could also fear that engineers could create systems they wrongly think have consciousness, thus disregarding what is really special about us, genuine conscious beings. So, given our mystery intuitions, sentientism might generate an attitude of humility directed at sentient creatures (whom we are not sure we could easily understand or replace) and potentially sentient ones (who might matter more than we know).

Sentientism encourages this attitude of humility, but it is neither necessary nor sufficient for it. Still, illusionists should keep in mind the value of this attitude. They can do it while adopting an objective view: thinking about various beings able to attain different goods, which might remain inaccessible to us, could give us a similar sense of humility. They can do it while adopting a preference view: thinking about the fact that other beings might care deeply about things which do not matter to us, and reminding ourselves that their care is as significant as ours, might also make us humble.

Here again, the religious analogy helps. A devout person can stress that the belief in God is an antidote to pride. Contemplating God makes us humble. An atheist who fears pride could concede this but insist that atheists, too, have ways to remain humble. They can remind themselves of the immensity of the universe, of the vastness of their ignorance, of the ephemeral or illusory character of their selves, of the meaninglessness of most of their pursuits, etc. Humility is not the believer's privilege. It does not have to be the sentientist's privilege either.

* * *

Illusionism departs from our intuitive ethical picture when it comes to the ground of value and moral status, but whether it does so about their distribution depends on the non-sentientist ethical view we adopt. I presented four such views. Some of them are plausible, and, combined with illusionism, they can account for ethical data. Ethical objections to illusionism can be answered.

[95] Schwitzgebel and Garza (2015); Metzinger (2021); Saad and Bradley (2022).

10

Shedding Our Cartesian Scales

Illusionism and Epistemology

Phenomenal consciousness seems epistemically significant. Does denying its existence have epistemological implications? If so, which ones? This chapter answers these questions. It focuses on illusionism's implications for epistemic justification and sets aside notions like knowledge or understanding. It structurally mirrors the previous chapter and therefore moves faster to avoid repetition.

First, I motivate the exploration of illusionism's epistemological implications. I focus on our intuitive epistemological picture. I show that, compared to this intuitive picture, illusionism is revisionary about the grounds of justification, but does not have to be massively revisionary about its distribution. I sketch three non-phenomenalist views of justification available to the illusionist, and close the chapter with concluding remarks.

10.1. Motivating the exploration

Philosophical zombies behave like phenomenally conscious beings. They say, "Here's an apple" when shown one, and "I could be fooled by an evil demon, but at least I know for sure that I am having an experience of an apple," when shown an apple during the seminar on Descartes. However, only phenomenally conscious beings are ever phenomenally presented with the apple. Only they have experiences that are directly introspectively presented to them. Intuitively, this is important to justify these beings' beliefs about apples and apple experiences. Zombies, lacking phenomenal experiences, seem to lack some of the justifications conscious beings have.

If illusionism is true, we are not different from zombies. This suggests that the view is epistemologically revisionary: we lack the justification we thought we had when we thought we were phenomenally conscious.

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0011

We should determine illusionism's epistemological implications for two reasons. First, if we have non-negligible credence in illusionism, we must do so to update our epistemological beliefs and practices accordingly.

Second, knowing illusionism's epistemological implications helps to assess the view, since, like in ethics, these implications motivate objections appealing to epistemological data about alleged manifest epistemological truths. One objection aims at directly refuting illusionism. It contends that illusionism entails the negation of manifest epistemological truths and can be ruled out. A more modest objection is that illusionism does not contradict epistemological data, but struggles to account for it, so that we can draw an inference to the best explanation of epistemological data against illusionism.

For instance, it has been claimed that, if illusionism is true, we are not justified in ruling out absolute nihilism, the view that *nothing* exists at all.[1] Since we are manifestly justified in doing so, illusionism is false. It has also been claimed that, if illusionism is true, we do not have immediate (non-inferential) *infallible* justifications.[2] As it is manifest that we have those, we can rule out illusionism. Assessing these objections requires clarifying illusionism's epistemological implications.[3]

I will make two presuppositions. First, we share an intuitive picture of justification. We tend to see phenomenality as real and as playing a role in grounding justification. We tend to agree on broad features of the distribution of justification: e.g., we have immediate justification to believe things about our perceptible environment, our minds, and our phenomenal experiences. Assuming the existence of this intuitive picture, we can put illusionism's implications in terms of revisions.

Second, I grant that, plausibly, some intuitive epistemological claims are manifest. They possess strong pre-theoretical justification and can legitimately count as epistemological data, at least for now. "We have immediate justification to believe various things about our perceptible environment" counts. For now, I suspend judgment about whether "we have some infallible immediate justification" and "we can conclusive rule out absolute nihilism" also count.

[1] C. D. Brown (2022).

[2] Merlo (2020).

[3] Other anti-illusionist arguments focus on illusionism's incapacity to account for some epistemological data that do not concern justification. Duncan (2023) argues that illusionism does not account for our knowledge of things (knowledge-of)—as opposed to propositional knowledge (knowledge-that). I see this argument as a variation on the phenomenal representation objection (Levine 2001, 143–49; Robinson 2022), already discussed in Chapter 7.

10.2. The intuitive character of epistemological phenomenalism

A. Varieties of epistemological phenomenalism

Our intuitive epistemological picture relies on phenomenal realism and epistemological phenomenalism. It views the phenomenal as real and as playing a role for justification. Here is one version of epistemological phenomenalism.

> *Weak Phenomenalism*: Some standard phenomenal states ground notable justification.

"Justification" means *propositional* justification. We have propositional justification when we have some pro tanto reason to believe a proposition—when something weighs in favor of believing it, so that, to a certain extent, it should be believed, whether or not we believe it. I set aside *doxastic* justification—when we really hold the belief on the basis of appropriate propositional justification. Propositional justification is closely related to what I called earlier *evidence* and *epistemic appearance*; each term corresponds to a given context of use. Justification is *notable* when it is non-negligible. *Standard* phenomenal states are those that realists think frequently occur. "Ground" means "ground or would ground": an illusionist can accept *Weak Phenomenalism*.

Weak phenomenalism is intuitive. Consider this.

> *Eve's Apple*: Eve wakes up. Before she remembers anything, she has a phenomenal visual experience of a red apple in front of her.

Intuitively, Eve has a justification to believe that there is a red apple in front of her, and a justification to believe that she has a phenomenal experience of a red apple. Intuitively, she has these justifications in virtue of her phenomenal experiences. Eve could be in a normal situation, waking up in her bedroom with an apple on her night table. She could also have been kidnapped during her sleep, put in a vat, and her brain connected to a computer feeding it with fallacious perceptual input. She could even have been spending her whole life as a brain in a vat or fooled by an evil demon. Still, in all these cases, having the right phenomenal experiences seems enough to give her some sort of pro tanto justification. This implies weak phenomenalism.

Considering *Eve's Apple* generates intuitions beyond weak phenomenalism. Eve's justifications seem immediate: they do not seem to depend on inferences from other beliefs. Eve's phenomenal experiences also seem to have peculiar justificatory properties: they seem introspectively obvious—that is, self-revealing and infallibly introspectively presented.

Consider a stronger form of phenomenalism:

> *Distinctive Phenomenalism*: Some standard phenomenal states ground notable justification, and they ground distinctive justification compared to other states.

Distinctive justification is justification much more significant than (or very different from) other justifications. It could be immediate, or self-revealing, or infallible justifications, for instance, if other states do not ground them. "Other states" means non-phenomenal states (among those we can positively conceive) otherwise close to phenomenal states, e.g., from a functional point of view. Zombie states are typical examples of such other states.

Finally, here is an even stronger form of phenomenalism, which entails distinctive phenomenalism, but not conversely.

> *Exclusive Phenomenalism*: Some standard phenomenal states ground notable justification, and nothing non-phenomenal grounds justification in reality.

Distinctive and exclusive phenomenalism both seem intuitive when we consider zombies.

- *Zombie Eve's Apple*: Zombie Eve wakes up. Before she remembers anything, she enters a non-phenomenal state physically and functionally similar to the state of Eve when she sees a red apple.

Zombie Eve tends to believe that there is an apple in front of her and that she phenomenally experiences it. Does she have a justification to believe so? Neither the apple nor the experience is phenomenally given to her. Some will intuit that she has no justification to believe this, nor anything else. This leads to exclusive phenomenalism.

Others will admit that she has no justification to believe in the apple or the apple experience, but intuit that she has some other justifications. She

might have justification to believe logical truths: for this, non-phenomenal reasoning seems enough. Zombie Eve's justification would still be very limited, and distinctive phenomenalism would remain attractive.

One might insist that Zombie Eve has immediate justification to believe things about her environment. Her tendency to believe that there is an apple is produced by a reliable belief-formation process. This might be enough to give her corresponding immediate justification. Still, this sort of consideration, which I discuss later, implies differences between the justifications had by Eve and Zombie Eve: she lacks justifications to believe that she has a phenomenal experience of an apple, since this belief of hers is unreliably produced. This still leads to distinctive phenomenalism.

Finally, one could feel that, as Zombie Eve is *disposed* to believe she has a phenomenal experience, she has a justification to believe so. I discuss this view later. Even in this case, Zombie Eve's justifications intuitively differ from Eve's: unlike Eve's, Zombie Eve's mental states are not self-revealing nor introspectively infallibly presented. Distinctive phenomenalism might be hard to avoid.

To conclude, weak phenomenalism seems strongly intuitive. Considering zombies also yield strong intuitive support for distinctive phenomenalism, and maybe also for exclusive phenomenalism.

B. Epistemological phenomenalism and blindsight

We do not need zombies to make distinctive phenomenalism intuitive. *Blindsight* is a well-studied condition in which, following lesions in the primary visual cortex, patients report a lack of visual experiences in a part of their visual field.[4] The mainstream view is that they lack these experiences. Yet, forced to choose, they make accurate guesses about visual stimuli in their blind areas—e.g., regarding the presence, position, movement, and orientation of visual items, as well as shapes, colors, or facial expressions.[5]

Intuitively, when blindsighters detect a circle in their blind field, they are not justified to believe that there is a circle. From their own perspective, they are just guessing. Experienced blindsighters might have *inferential* reasons

[4] For discussions of phenomenalism through the lens of blindsight, see notably Smithies (2019); Silins (2024).

[5] Weiskrantz (2007).

to endorse their guesses, as they know that past guesses were accurate. Still, focus on *immediate* justification: in this respect, blindsighters seem quite different from normal subjects. Intuitively, a phenomenal experience of a circle gives immediate justification to believe that there is a circle: it makes the circle appear-as-present to the subject. A blind detection of the circle does not. Phenomenal experiences seem to have distinctive justificatory powers.

Blindsight differs functionally from normal vision. Could aspects of this functional difference, rather than the phenomenal difference, ground the epistemological difference?

First, blindsight patients do not make their guesses spontaneously. They do not tend to believe their guesses. However, following Ned Block, imagine an upgraded form of blindsight: *super-blindsight*.[6] Super-blindsighters form guesses spontaneously and confidently believe them, while still correctly reporting the absence of related experiences. Super-blindsight might be impossible, but it seems imaginable. Intuitively, even a super-blindsighter lacks immediate justification to believe there is a circle when they guess and believe there is one. The genuine presentation involved in a phenomenal experience of a circle seems to be missing.

Imagine that blindsight's discriminatory powers increase to match those of phenomenally conscious vision: blindsighters now also form confident and accurate metacognitive beliefs about their own states, thus becoming *super-duper-blindsighters*.[7] For some, even super-duper-blindsight intuitively lacks the justificatory powers of phenomenally conscious vision.[8] Even super-duper-blindsight does not involve the genuine presentation at play in phenomenal experiences. This tentatively supports distinctive phenomenalism. Super-duper-blindsight states are functionally close to phenomenal experiences. They lead reliably and spontaneously to confident beliefs—first-order and higher-order. Still, intuitively, even they fail to provide the kind of immediate justification that phenomenal experiences give.

While this makes distinctive phenomenalism attractive, it is coherent to deny distinctive phenomenalism but concede that super-duper-blindsight is epistemologically unlike phenomenal vision. Even super-duper-blindsight states are *not* zombie states: contrary to zombie states, they do not trigger phenomenal introspection.

[6] See Block (1995, 233) for the idea. My characterization of super-blindsight follows Smithies (2019, 85–86).

[7] The expression is due to Block, but the insistence on metacognition comes from Smithies.

[8] Smithies (2019, 88–90).

10.3. The wide acceptance of epistemological phenomenalism

Epistemological phenomenalism was historically popular. Classical rationalists (Descartes, Malebranche, Spinoza, etc.) were probably committed to exclusive phenomenalism. They saw knowledge, and, arguably, justification, as requiring clear and distinct ideas, naturally understood as certain phenomenal experiences. Classical empiricists (Locke, Hume, etc.) saw all knowledge as relying on experience, which also arguably entails exclusive phenomenalism.

These views were versions of *foundationalism*, on which all justifications ultimately depend, by way of actual or possible inferences, on immediate justification. Historically, foundationalism was "very widely, almost universally accepted."[9] Traditional foundationalists were also internalists.[10] They saw justification as depending on factors internal to the subject: mental features, or necessarily accessible features. Phenomenal experiences plausibly count as internal in both senses.

Other historically influential versions of internalist foundationalism entailed exclusive phenomenalism. This is the case of Russell's view,[11] on which all knowledge relies on *acquaintance*—acquaintance being naturally interpreted as involving phenomenal experiences. Some recent foundationalists maintained similar ideas, entailing exclusive phenomenalism.[12]

Recent influential views of justification, called "phenomenal conservatism,"[13] or "dogmatism,"[14] partly uphold the spirit of internalist foundationalism. They contend that we have *some* internally determined immediate justification, without necessarily claiming that *all* justification depends on it. On these views, perceptual phenomenal experiences are epistemologically crucial. They possess "phenomenal force,"[15] and *present to us how things are*, thus providing internally determined immediate justification. So, on these views, some phenomenal experiences—perceptual ones—play a distinctive role for justification,[16] which implies distinctive

[9] Hasan and Fumerton (2022).
[10] Pappas (2014).
[11] Russell (1912).
[12] Fumerton (1995; 2016); BonJour (2000).
[13] Huemer (2001; 2007).
[14] Pryor (2000). For related views, see Brogaard (2018); Smithies (2019).
[15] Pryor (2000, n. 37).
[16] In some views, non-perceptual experiences also play an important role (Chudnoff 2012; 2014; Kriegel 2023).

phenomenalism. On some of these views,[17] phenomenal experiences are also crucial to inferential and mediate justification, which naturally leads to exclusive phenomenalism.

Epistemological phenomenalism is influential, but debated. Radical coherentist and anti-foundationalist views, on which there is no immediate justification—the *coherence* of our beliefs with each other being what justifies them—do not give a particular role to phenomenal experiences.[18] However, that these views have this consequence is often seen as one of their weaknesses. Moderate coherentism, on which coherence plays *some* justificatory role, is compatible with experiences playing a distinctive role for justification.[19]

One can deny experiences' epistemological peculiarity but admit immediate justification, as well as internalism and foundationalism. *Beliefs*, rather than experiences, could be foundational. This is *doxastic conservatism*:[20] the view that merely having a belief gives some justification for the proposition believed. Still, the most popular alternative to epistemological phenomenalism is a form of *externalist* foundationalism: reliabilism.[21] On a standard version of reliabilism, we have immediate justification for some propositions we believe, in virtue of factors which are not necessarily internal and concern the fact that the beliefs were produced by *reliable* processes—i.e., processes likely to produce true beliefs. Mediate justification happens when we form beliefs on the basis of a conditionally reliable process that takes justified beliefs as inputs.[22]

Reliabilism is influential.[23] Many find it promising, notably because reliability lends itself to a naturalistic account. Reliabilism does not give a distinctive role to phenomenal experiences, and naturally opposes distinctive and exclusive phenomenalism. Still, as some reliabilists noted, the "nonnecessity of experience" for justification is an implication of reliabilism that is "rarely or never made explicit," and which is itself responsible for a "great deal of resistance to reliabilism."[24] This confirms that distinctive phenomenalism,

[17] For instance, Huemer (2013).
[18] Olsson (2023).
[19] Pryor (2000, 535).
[20] Vahid (2024).
[21] Goldman (1979); J. Lyons (2009); Goldman and Beddor (2021).
[22] Reliabilism is standardly presented as a theory of *doxastic* justification. I am assuming here that it can be reinterpreted as a view of propositional justification.
[23] According to the 2020 PhilPapers survey, reliabilism is the most popular view of justification among contemporary philosophers, with 34% of respondents accepting it or leaning toward it (Chalmers and Bourget 2020).
[24] J. Lyons (2009, viii).

at least, is highly intuitive. Reliabilism is technically compatible with weak phenomenalism, since some phenomenal experiences could constitute reliable belief-formation processes, but phenomenal factors alone are unlikely to ground the reliability of a process.[25]

So, many epistemological views imply at least distinctive phenomenalism. Some views, like reliabilism, imply that distinctive phenomenalism is false, but this implication is among their counterintuitive consequences.

Some have noted that phenomenalism clashes with some views of consciousness. It has been argued that, given various scientific theories of consciousness, phenomenal consciousness is unlikely to be necessary for justification,[26] which contradicts exclusive phenomenalism. It was also argued that, if reductive materialism is true, phenomenal consciousness is not 'naturally' distinctive. It does not have highly distinctive causal and explanatory features, which in turn suggests that it is not epistemologically distinctive compared to non-phenomenal processes with otherwise comparable profiles,[27] which enters in tension with distinctive phenomenalism. Yet another argument is that at least one scientific view of consciousness (HOT) implies that phenomenal consciousness is insufficient for justification,[28] implying that weak phenomenalism is false. Of course, all this is compatible with all varieties of phenomenalism being intuitive. In what follows, "phenomenalism," unqualified, means distinctive phenomenalism.

10.4. Assessing illusionism's epistemological consequences

Phenomenal realism is intuitive and widely accepted. The same is true of epistemological phenomenalism. Our intuitive epistemological picture is one in which the phenomenal is real and plays a distinctive role in grounding justification.

If so, illusionism is revisionary about the grounds of justification, which cannot be as our intuitive picture depicts it. Does it imply revisions for the *distribution* of justification? As previously with ethics, this depends on our answer to two key questions. First, which link between phenomenality

[25] An other influential family of views departing from epistemological phenomenalism is virtue epistemology (Sosa 1991; Turri et al. 2021), which I leave aside here.
[26] Berger et al. (2018).
[27] G. Lee (2014).
[28] Berger (2020).

and justification do we admit? Second, what, aside from phenomenal experiences, grounds justification?

Start with the first question. Illusionists who concede *exclusive* phenomenalism conclude that we do not have justification to believe anything. This is epistemological nihilism: a radically revisionary, and probably unacceptable view. Illusionists who concede only *distinctive* phenomenalism are led to significant revisions. They can contend that we have justifications, but concede that they are quite distinct from those our intuitive picture admits.

Illusionists conceding merely weak phenomenalism can technically avoid distributional revisions, but, like in ethics, this might lead them to admit that the justifications that would be had by phenomenally conscious beings would have *two* distinct full grounds—an amazing coincidence. They are also not guaranteed against distributional revisions, and neither are illusionists rejecting even weak phenomenalism. Whether they are led to revisions depends on their answer to the second key question.

Now, let us address the first question. Which link between phenomenality and justification should we admit?

Weak phenomenalism is supported by a strong and conceptual intuition, stemming from the way phenomenal concepts characterize phenomenal experiences, as direct evidential states providing *direct evidence* (i.e., *immediate justification*) to the subject. Perceptual experiences, for instance, are grasped as providing direct evidence for the instantiation of sensory properties. Like previously with weak *ethical* sentientism, I doubt that our intuition for weak phenomenalism can change, and I doubt that we can find good reasons to discard it. Illusionists should concede weak phenomenalism, as an innocuous conceptual truth concerning non-existent entities. Note that the intuition that phenomenal experiences are introspectively obvious—self-revealing and introspectively infallibly presented—is also conceptual. I will come back to this later.

Now, consider distinctive and exclusive phenomenalism. They are intuitive, but like in ethics, we should be suspicious of our intuitions. Distinctive and exclusive phenomenalism do not just assert the justificatory powers of phenomenal states. They compare them with those of non-phenomenal states, like zombie states, and our intuitions about the justificatory powers of non-phenomenal states should be treated with caution. We are intuitive phenomenal realists and falsely tend to think, for instance, that Zombie Eve lacks something we have and which we think of as a source of justification.

With status quo bias kicking in, this easily explains why we would think that the justifications she has—if any—are distinctively less significant than ours. On this explanation, the resulting intuitions are not conceptual, but stem from false presuppositions, and are cognitively penetrable. Like previously with ethical sentientism, phenomenal realism and status quo bias being intuitively compelling would explain why these intuitions about zombie states' relative lack of justificatory powers would behave like compelling intuitions: illusionists too could feel that Zombie Eve lacks something important when it comes to justification. If this explanation of the intuition is correct, understanding it undermines its evidential status.

What about the intuitive idea that blindsight, even in upgraded forms, lacks the justificatory powers of phenomenally conscious vision? It does not exactly entail distinctive phenomenalism, but it fits well with it. I discuss it later.

For now, admit that distinctive and exclusive phenomenalism are intuitive. Assuming the explanation of these intuitions I gave, their evidential value is undermined. Should we reject these forms of phenomenalism? This depends on how we answer our second key question.

I will show that there are coherent and attractive non-phenomenalist views of justification. Rejecting exclusive phenomenalism seems natural. Still, these non-phenomenalist epistemological views, combined with illusionism, lead to admitting that the justifications we have are unlike those grounded by the phenomenal in our intuitive picture. The phenomenal seems obvious, in virtue of conceptual intuitions, and our best non-phenomenalist epistemological views imply that nothing is obvious in the same way. We should concede a form of distinctive phenomenalism and admit related revisions.

10.5. Alternatives to epistemological phenomenalism

What, aside from phenomenal experiences, grounds justification? I will examine three non-phenomenalist views of justification: epistemological quasi-phenomenalism, reliabilism, and epistemological psychologism. These do not exhaust the possibilities, but hopefully, they capture most attractive ones. Since, in our intuitive picture, the specific justificatory powers of phenomenal experiences notably concern immediate justification, I focus on immediate justification.

A. Quasi-phenomenalism

For epistemological quasi-phenomenalism, quasi-phenomenality roughly plays the role played by phenomenality in epistemological phenomenalism. Quasi-phenomenal states play a distinctive (if not an exclusive) role for epistemic justification. For instance, my quasi-phenomenal perception of an apple gives me immediate justification to believe that there is an apple and that I have an apple experience.

Exclusive quasi-phenomenalists add that nothing else provides justification. Distinctive quasi-phenomenalists simply claim that nothing else provides comparable justification: e.g., nothing else might provide immediate justification, even if non-phenomenal factors, like the coherence of our beliefs, can play a role in inferential justification.

Quasi-phenomenalism implies that we have plenty of justifications, for instance, regarding our perceptible environment. It partially upholds the spirit of our intuitive picture and minimizes distributional revisions: at least in the human case, quasi-phenomenality is present where we thought phenomenality was present. It implies that blindsight, even in upgraded forms, is not justificatory like normal vision, which involves quasi-phenomenal states. Even super-duper-blindsight states are not quasi-phenomenal states: if they were, they would normally trigger phenomenal introspection.

Still, quasi-phenomenalism does not leave our intuitive picture untouched. Quasi-phenomenal states do not play exactly the same justificatory role as phenomenal experiences in our intuitive picture. They are not self-revealing: we have them without having immediate justification to believe that they have the nature they really have. If anything, they give us immediate justification to have *false beliefs* about their nature. They are also not introspectively infallibly presented: to the extent that they are introspectively presented, they are even presented fallaciously, as phenomenal.

Quasi-phenomenalism faces the same two challenges as ethical quasi-sentientism, concerning *determinacy* and *justification*.

First, *determinacy*. If the distribution and extension of quasi-phenomenality is, as I argued, probably widely indeterminate, then, given quasi-phenomenalism, so is epistemic justification. This is an uncomfortable consequence. The dialectic resembles the one developed in the previous chapter. The determinacy issue is a problem, but maybe not a fatal one, since

the quasi-phenomenalist can hope that quasi-phenomenality will turn out to be mostly determinate. The determinacy issue seems also less pressing in epistemology than in ethics: quasi-phenomenality is most likely to be indeterminate in non-humans, but it is plausible that, in epistemology, we care mostly about humans.

Second, *justification*. Why should we adopt epistemological quasi-phenomenalism? We might have reasons to accept epistemological *phenomenalism*, since the idea that phenomenal experiences play some epistemological role is intuitive and follows from how we represent them, but our concepts of quasi-phenomenal states do not characterize them as justificatory. The view that they have distinctive justificatory powers seems unmotivated. This is manifest with blindsight: quasi-phenomenalism *accommodates* our intuition that all varieties of blindsight are not as justificatory as normal vision, but it does not give us reasons to believe this intuition is correct.

Here again, the dialectic mimics the dialectic of the previous chapter. Quasi-phenomenalists can say that *introspection*, not conceptual reflection, tells us that some of our mental states, which turn out to be quasi-phenomenal, are distinctively justificatory. However, illusionism implies that introspection is mischaracterizing and thus unreliable when it ascribes special properties. Even if introspection tells us that quasi-phenomenality is distinctively justificatory, it is not to be trusted in this respect.

The epistemological quasi-phenomenalist might have an answer to the justification challenge that was unavailable to the ethical quasi-sentientist. *If* quasi-phenomenalism is correct, quasi-phenomenality *really* has certain distinctive justificatory powers, and it is a small leap to claim that quasi-phenomenal states also justify our beliefs about their justificatory powers. The result is coherent, but this bootstrapping, reminiscent of externalist responses to skepticism, seems cheap. Why believe quasi-phenomenalism in the first place, if we are not already convinced of its truth? And arguably, we are not, since our intuitive starting point is phenomenalist.

Finally, epistemological quasi-phenomenalism faces the specific problem of *Atypical Illusions of Phenomenality*. Quasi-phenomenal experiences are states that typically trigger phenomenal introspection in us, but phenomenal introspection could also be triggered atypically, by things *other* than quasi-phenomenal states. Suppose a mad neurosurgeon makes me directly introspect a phenomenal experience of red in the absence of a quasi-phenomenal experience of red. Intuitively, in this case, I have immediate justification to

believe that I have a phenomenal experience of red, since I introspect it, but quasi-phenomenalism apparently implies that I do not.

B. Reliabilism

Reliabilism, which was mentioned earlier, states that what matters for justification is the reliability of belief-forming processes: reliability *tout court* for immediate justification, conditional reliability with justified beliefs being used as inputs for inferential justification. Here, I focus on immediate justification.

Reliabilism can be exclusive (if reliability is all that matters for justification) or distinctive (if it distinctively matters). Both imply that we can have immediate justifications without phenomenal experiences, as long as the relevant beliefs are appropriately formed. For reliabilists, "zombies have no [phenomenal] sensations at all, yet they have justified perceptual beliefs."[29]

Given certain strong assumptions, reliabilism and quasi-phenomenalism are equivalent. Suppose my Meta-Cartesian view is right. Phenomenal introspection tracks our non-inferential dispositions to believe, which they mischaracterize as phenomenal. However, the reason why our phenomenal introspection tracks non-inferential dispositions to believe is that these dispositions signal reliable non-inferential belief-formation processes. So, *being a reliable non-inferential belief-formation process* is a typical property of quasi-phenomenal processes, which phenomenal introspection has ultimately the function to track. Now, suppose that, as I argued earlier is unlikely, one property of our quasi-phenomenal states is a uniquely salient and important natural kind, to which our concept of quasi-phenomenality determinately refers. Suppose that this property is precisely the property of being a reliable non-inferential belief-formation process. On these assumptions, quasi-phenomenal states are nothing but reliable non-inferential belief-formation processes. Quasi-phenomenalism is a version of reliabilism.

These assumptions are costly. Plausibly, some reliable non-inferential belief-formation processes fail to be quasi-phenomenal. Still, reliabilism and quasi-phenomenalism remain close: what plays a distinctive role for immediate justification, for a reliabilist, is a property that all (or most)

[29] J. Lyons (2009, 54).

quasi-phenomenal states have. Since the reverse is less plausible, reliabilism probably ascribes justification more liberally than quasi-phenomenalism.

Supposing that phenomenal introspection has the function to track reliable non-inferential dispositions to believe, reliabilists can say that the real source of immediate justification—reliable non-inferential belief-formation processes—is what our evidential cognition and phenomenal introspection are tracking, and that they represent correctly some of their evidential properties, although they also caricature these properties by mischaracterizing the processes as phenomenal. This is an elegant result. It makes the view attractive, given our phenomenalist starting point.

Reliabilism opposes phenomenalism: it is revisionary about the grounds of justification. Is it also revisionary about its distribution?

Reliabilism implies that we have justifications to believe many things, e.g., about our perceptible environment. Still, in specific cases, it leads to counter-intuitive consequences. Many remarked that it ascribes justification very liberally. It counterintuitively implies that a person with clairvoyant powers, who lacks evidence for or against having these powers, and whose powers directly cause a belief, has a justified belief.[30] This is the clairvoyance problem.

Others stressed that reliabilism is too restrictive. Suppose a world in which people have perceptual beliefs similar to ours, but caused by fallacious perceptions generated by an evil demon. Intuitively, they have false but *justified* perceptual beliefs, but for reliabilism, their perceptual beliefs, which are unreliably caused, are not justified. This is the new evil demon problem.[31]

Reliabilists have answers to these problems,[32] which nevertheless suggest that reliabilism is revisionary when it comes to the distribution of justification.[33] A reliabilist could contend that these revisions are too modest or that they concern cases too marginal to count as contradicting epistemological data.

Aside from this, illusionist reliabilists, similar to quasi-phenomenalists, must probably admit that none of our mental states are self-revealing or infallibly introspectively presented. Our quasi-phenomenal states are neither

[30] BonJour (2000).
[31] Lehrer and S. Cohen (1983); S. Cohen (1984).
[32] Goldman and Beddor (2021).
[33] In spite of the claims of some reliabilists that the view captures our "ordinary conception" of justification (Goldman 1979, 11): my presupposition that our intuitive picture of justification is phenomenalist is at odds with this reliabilist claim. The idea that reliabilism (about knowledge, but the point could be extended to reliabilism about justification) captures our folk epistemology has been challenged empirically (Turri 2016).

self-revealing nor infallibly introspectively presented. For reliabilists, the same is probably true of all our mental states. Some revisions seem unavoidable.

Our intuitions about blindsight, easily accommodated by quasi-phenomenalists, challenge reliabilists. They can explain that ordinary blindsight does not provide justification, since it does not generate beliefs, but they must probably concede that super-blindsight is justificatory, which is counterintuitive.[34] This difficulty resembles the clairvoyance problem.

To temper the issue, reliabilists can stress that there is an important epistemological difference between even super-blindsight and normal perception. Both provide justification for first-order beliefs regarding a stimulus, but only the second is introspected: only about the second do we reliably form metacognitive beliefs. So, while super-blindsight and normal vision both give first-order justification to believe things about stimuli, only the second gives justification for metacognitive beliefs.

A difference subsists even between *super-duper*-blindsight, which triggers metacognitive beliefs, and normal vision. We can make this difference clear using the framework of Meta-Cartesianism: super-duper-blindsight triggers metacognitive beliefs, but not *evidential* self-beliefs regarding immediate justification, since it is not represented by our Cartesian Device responsible for representing direct evidence, that is, responsible for representing phenomenal properties. In super-duper-blindsight, we reliably form first-order beliefs about stimuli and metacognitive beliefs about our own states, but not evidential self-beliefs about our own immediate justifications. Otherwise, super-duper-blindsight would activate our Cartesian Device and be introspected as phenomenal. Hence, super-duper-blindsight gives first-order justifications about stimuli, metacognitive justifications about its own presence, but not evidential ones about its own justificatory powers. Reliabilists can maintain that quasi-phenomenal states have a sort of epistemological specificity in this restricted sense.

Illusionists reliabilists can also explain the counterintuitive character of their view's additive revisions about blindsight or clairvoyance in a way that makes this character harmless. Indeed, since blindsight fails to trigger phenomenal introspection, it is not represented as direct evidence by our Cartesian Device: intuitively, it does not provide immediate justification. Still, illusionists have no reason to believe that our intuitive representations

[34] But see J. Lyons (2009, 52–54).

of direct evidence are infallible—quite the contrary. We should expect the correct theory of justification—supposedly, reliabilism—to contradict such intuitions about which mental states give immediate justification. The idea that evidential cognition makes mistakes is natural given Meta-Cartesianism: If evidential cognition, relying on our Cartesian Device, deeply mischaracterizes its targets, why not admit that it also makes other mistakes, like missing the justificatory powers of capacities such as blindsight?

Illusionists endorsing reliabilism face a specific challenge, concerning the *Prima Facie Justifiability of Realism*. Suppose illusionism is true. Gottfried, unaware of the illusionist arguments, thinks, on the basis of introspection, that he has phenomenal experiences. Intuitively, he is justified in doing so, but reliabilism cannot explain this, since phenomenal introspection is *unreliable*. Quasi-phenomenalism did not encounter this problem, but only a more specific one linked to *Atypical Illusions of Phenomenality*. This new problem is more serious, as it concerns a familiar case. It also raises a dialectical problem for illusionists: as noted previously, illusionism is dialectically stronger if it accounts for the prima facie justifiability of realism.

Illusionist reliabilists can answer that, while *phenomenal introspection* is unreliable, *introspection in general* is reliable, which makes Gottfried's false belief justified. This hints at the "generality problem" for reliabilists, which is independent of illusionism.[35] Token belief-formation processes belong to various types. Some of them are unreliable (e.g., phenomenal introspection) while others are reliable (e.g., introspection). The generality problem is the problem of determining which type matters for the justification of a given belief.[36] Ideally, an illusionist reliabilist should solve the generality problem in a way that accounts for the prima facie justifiability of realism. This is no trivial task, and the existence of the generality problem raises the specter of indeterminacy for reliabilism.

Alternatively, illusionist reliabilists can bite the bullet and grant that Gottfried's beliefs are unjustified. They should at least explain our opposite intuition. Indexical forms of reliabilism, on which "reliable" is correctly applied to processes reliable *in our world*, can help.[37] These views supposedly

[35] Goldman and Beddor (2021).

[36] A reliabilist admitting that sensory perception mischaracterizes sensible qualities as primitive qualities faces a similar problem when accounting for the justification of some of our perceptual beliefs.

[37] Comesaña (2002). Thanks to Keith Harris for the suggestion.

solve the new evil demon problem, by entailing that beliefs in the demon world are reliable (and thus justified) since they are formed through processes (e.g., *perceiving*) reliable in our world. They can also explain our intuitions about Gottfried, whose phenomenal beliefs are false, but formed through a process, phenomenal introspection, which we tend to think of as reliable in our world—since we are intuitive phenomenal realists. Indexical reliabilism, given our strong realist intuitions, explains why we intuit that Gottfried's beliefs are justified, even if they are not. One other strategy for reliabilists, inspired by other responses to the new evil demon problem, states that Gottfried's beliefs are not justified, but that they have some other desirable epistemic quality, such as *blamelessness*, thus supposedly accounting for our false impression that they really are justified.[38]

C. Epistemological psychologism

For *epistemological psychologism*, sufficiently strong non-inferential dispositions to believe ground immediate justification.[39] In other words, for epistemological psychologism, *psychological appearances* ground *epistemic appearances*. As I understand it, epistemological psychologism is neutral on inferential justification, which could be grounded like immediate justification, in appropriately strong *inferential* dispositions to believe, or in a reliabilist fashion, etc.

Distinctive epistemological psychologism states that this immediate justification is distinctive. *Exclusive* epistemological psychologism states that nothing else grounds immediate justification. Here, I focus on distinctive epistemological psychologism.

Given certain strong assumptions, epistemological psychologism is a version of quasi-phenomenalism. If Meta-Cartesianism is right, phenomenal introspection tracks our non-inferential dispositions to believe, which they mischaracterize as phenomenal. *Being a certain sufficiently strong non-inferential disposition to believe* is among the properties of our quasi-phenomenal states. Now, suppose that, as I argued is unlikely, one property of our quasi-phenomenal states is a uniquely salient and important

[38] This distinction between justifiedness and blamelessness has been the focus of attention recently (Boult 2017; Pritchard 2012).

[39] Epistemological psychologism can be seen as a variation on doxastic conservatism, mentioned above. See Swinburne (2001, 140–42) for a view of the sort.

natural kind, to which our concept of quasi-phenomenality determinately refers, and suppose that it is precisely the property of being a certain sufficiently strong non-inferential disposition to believe. On these assumptions, quasi-phenomenal states are nothing but the right non-inferential dispositions to believe. Epistemological psychologism is equivalent to quasi-phenomenalism.

These assumptions are strong. Plausibly, some non-inferential dispositions to believe are *not* quasi-phenomenal, even if our quasi-phenomenal states typically are non-inferential dispositions to believe. We should expect psychologism, like reliabilism, to ascribe justification more liberally than quasi-phenomenalism.

If, as I supposed earlier, sufficiently strong non-inferential dispositions to believe typically involve reliable belief-formation processes, reliabilism and psychologism will often make convergent verdicts. Both account for our having justifications to believe various things about our perceptible environment. Still, they diverge in some cases: as some of our strong non-inferential dispositions to believe probably do not result from reliable processes, psychologism will be more liberal than reliabilism.

Given psychologism, *super-blindsight* is justificatory. Epistemological psychologists can attenuate the challenge this raises, like reliabilists. First, they can say that normal vision, involving quasi-phenomenal states, is still epistemologically distinct, since it is only about normal vision that we tend to form evidential beliefs. Therefore, we will have second-order justification about our first-order justification when engaged in normal vision, not in blindsight. Second, to the extent that seeing blindsight as justificatory is counterintuitive, they can explain our incorrect intuition: intuitions about which states have justificatory powers are downstream of our evidential cognition, which, according to Meta-Cartesianism, does *not* represent blindsight states as providing direct evidence—otherwise, blindsight states would seem phenomenal.

Epistemological psychologism might seem too liberal when considering subjects with exotic doxastic tendencies, like Capgras subjects who think their spouse has been replaced by a lookalike. Should we concede that Capgras patients are justified to believe in their delusion, as psychologism (vs. reliabilism) seems to says? This seems implausible. However, defenders of psychologism can stress that the justification is defeasible: Capgras patients' delusional beliefs might be prima facie justified, but they are not ultima facie justified.

Psychologism might fare better than quasi-phenomenalism with indeterminacy issues, as its key factor plausibly corresponds to a single psychological kind—at least, if the inferential/non-inferential distinction is reasonably determinate. Degreed indeterminacy might be feared when it comes to how strong a non-inferential disposition to believe must be to be justificatory, but positing corresponding degrees of justification might be enough to avoid the difficulty.

Can psychologism be justified? Having a strong non-inferential disposition to believe something means that, whatever we otherwise believe, we tend to form the belief in a way that forcefully bears on our cognitive economy. *Forming indeed* the corresponding belief, when there are no considerations to the contrary, seems the most natural, and a minima the most excusable epistemic response. The intuitive (though debatable) link between epistemic justification and epistemic blamelessness can be leveraged to claim that the disposition grounds immediate justification. Like the reliabilist, the epistemological psychologist could also insist that what they see as the source of immediate justification—strong non-inferential dispositions to believe—is what our evidential cognition and phenomenal introspection try imperfectly to get at, making the view attractive given our phenomenalist starting point.

Proponents of phenomenalism have criticized the basic tenet of epistemological psychologism.[40] I will not examine the objections in detail. I concede that, intuitively, phenomenal experiences are more plausible sources of immediate justification than mere dispositions to believe. However, this is to be expected, since phenomenal introspection tracks certain dispositions to believe and grasps them, under an evidential guise, as phenomenal experiences, characterized as states grounding immediate justification. Our evidential cognition indeed represents immediate justification as grounded in phenomenal experiences, which is why the idea is so intuitive. Still, once we embrace illusionism, the view that the dispositions to believe are themselves justificatory becomes a contender.

Unlike reliabilism, epistemological psychologism easily deals with the *Prima Facie Justifiability of Realism*. Given psychologism, as we are strongly non-inferentially disposed to believe in phenomenal experiences, we are prima facie justified to believe in them, even if the process of belief-formation is unreliable. Psychologism also handles *Atypical Illusions of Phenomenality* better than quasi-phenomenalism. When phenomenal introspection is

[40] See Huemer (2007, 31), who notably ascribes a related view to Armstrong (1961, 84–87).

triggered atypically, we are strongly and non-inferentially disposed to believe we have a phenomenal experience. Given epistemological psychologism, we have the same prima facie immediate justification to believe that we have the experience as when we enter a quasi-phenomenal state triggering introspection 'normally'.

Finally, like reliabilism and quasi-phenomenalism, psychologism probably cannot avoid some distributional revisions. Our non-inferential dispositions to believe are not introspectively obvious: they are neither self-revealing nor infallibly introspectively presented.

D. Assessing alternatives to phenomenalism

I find reliabilism and epistemological psychologism more attractive than quasi-phenomenalism. Hybrid options are also available, which I have not discussed yet. The hybrid view I find the most interesting is pluralistic. It distinguishes between various concepts of justification,[41] JUSTIFICATION$_1$ and JUSTIFICATION$_2$, corresponding to distinct epistemologically desirable features: say, truth-conduciveness on the one hand, and blamelessness (which is indirectly connected to truth-conduciveness but distinct from it) on the other. Reliabilism could be correct of justification$_1$ and correctly captures truth-conduciveness, while epistemological psychologism could be true about justification$_2$, and correctly captures blamelessness.

I claimed earlier that, assuming that distinctive and exclusive phenomenalism are made intuitive because of weak phenomenalism being a conceptual truth, together with realist presuppositions and status quo bias, illusionism undermines the evidential value of this intuition. Since I just showed that there are plausible non-phenomenalist views of justification, exclusive phenomenalism appears extremely implausible. Illusionists do not have to become epistemological nihilists.

Distinctive phenomenalism is a different story. It remains true that, out of conceptual necessity, the phenomenal is justifying, as well as introspectively obvious. On the other hand, all non-phenomenalist views of justification I found attractive imply that nothing real plays this exact epistemological role: nothing real is introspectively obvious. This is hard to avoid given Meta-Cartesianism: we think of phenomenal experiences as entities playing

[41] In the spirit of Alson (2005).

a unique epistemological role, and if some actual mental states had the right profile of occurrences and played this epistemological role, it would be plausible to identify these mental states with phenomenal experiences, and illusionism, plausibly, would be false. Therefore, we should concede a form of distinctive phenomenalism: phenomenal experiences do not exist, but if they existed, they would ground a certain sort of justification, which, given our best views of what justification really is in our world, does not occur in reality.

10.6. Shedding our Cartesian scales

Take a step back. At an earlier evolutionary stage, we were disposed to believe various things, inferentially or non-inferentially: things *psychologically* appeared to us. These psychological appearances usually signaled reliable belief-formation processes since, plausibly, capacities generating these dispositions were selected because they were reliable.

We developed a mechanism tracking these dispositions and representing them as providing us with evidence, as making things *epistemically* appear to us: evidential cognition. This was advantageous. Our evidential cognition notably relied on a Cartesian Device representing direct evidential states—phenomenal experiences. These states were thus always also represented as phenomenal appearances: as states grounding specific epistemic appearances in a phenomenal medium, making other things directly appear to the subject, while themselves being introspectively obvious (self-revealing and infallibly directly appearing).

We became illusionists and understood that there are no phenomenal experiences: our evidential and phenomenal cognition (i.e., our Cartesian Device) mischaracterizes what happens in us. This does not imply that we do not have justifications to believe anything, that there is no evidence, or that there are epistemic appearances. The concept of a justification is separable from the concept of a phenomenal experience, even if the latter is the concept of a certain kind of justification. We can maintain that we have justifications, including immediate justifications, which are not grounded in phenomenal experiences.

We shed our Cartesian scales. We repudiate our intuitive phenomenalist understanding of justification, which we mobilized in past inquiries. Like the old skin of a snake, our innate evidential cognition served us well: thinking

about phenomenal experiences as our epistemic window on the world, we made sense of illusions, hallucinations, or false impressions. This tool took us, with Galileo and Descartes, all the way through the Scientific Revolution and Modern Philosophy. Now, it generates intractable problems when we try to fully integrate our minds within our scientific and physical understanding of reality. The problem evaporates as we admit that phenomenal introspection mischaracterizes both what happens in our mind *and* our epistemic justifications.[42]

Having shed our Cartesian scales, we are left with a choice between various theoretical possibilities regarding what grounds epistemic justification. I presented some of these possibilities. All of them are revisionary to some extent. In all of these theories, we have to admit that we have no actual mental states with the sort of introspective obviousness our phenomenal concepts ascribe to phenomenal experiences. In many of these views, processes like super-blindsight, counterintuitively, provide immediate justification.

Getting rid of our intuitive phenomenalist conception of justification and endorsing a new non-phenomenalist conception: both steps remain counterintuitive to some extent. Evidential and phenomenal cognition are determined by robust features of our innate Cartesian Device. Whatever theory we embrace, we will tend to think that we have phenomenal experiences that ground justifications. A dose of epistemological psychologism can explain why this belief is prima facie justified, or at least, blameless. Other views might explain at least why it seems so.

A few final clarifying remarks. First, throughout this book, and notably in this chapter, the same notions occurred repeatedly under various names. Epistemic *justifications* and *reasons to believe* are the same as *epistemic appearances* and *evidence*. *Direct* evidence is the same as *immediate* justification. I simply used "epistemic appearances" when determining how phenomenal consciousness appears to be, "justification" when discussing what we should do with this epistemological notion once we embrace illusionism, and "evidence" to discuss what our innate Cartesian Device represents.

Second, my treatment of the ethical and epistemological implications of illusionism differs. In ethics, I rejected distinctive *sentientism*, but I admitted a version of distinctive *phenomenalism* in epistemology: non-phenomenal

[42] The detour by illusionism is not needed to motivate giving up our intuitive understanding of justification. Naturalistic considerations already make reliabilism, for instance, attractive over phenomenalism. Reliabilism is compatible with both illusionism and realism, but it already departs from the intuitive picture of justification our innate evidential cognition suggests.

states cannot play the unique epistemological role phenomenal experiences intuitively play. This differential treatment is understandable if we remember that the link between phenomenal and evidential cognition is deep and direct: we think of phenomenal experiences *as certain evidential states*. On the other hand, the link between phenomenal cognition and our representations of value and moral status is indirect and mediated notably by affective cognition. This is why, embracing illusionism, we should concede that the justifications we actually have are different from those supposedly grounded in phenomenal experiences, while we do not have to do it for value and moral status. This is also why we should admit that real justification is often grounded in the actual states we misrepresent as phenomenal, while the same might not be as true in ethics. Indeed, the most attractive non-phenomenalist views of immediate justification give a key role to features like *being a non-inferential disposition to believe*, or *being a reliable non-inferential process*—which are features had by our quasi-phenomenal states, and which our evidential cognition might plausibly have, maybe among other features, the function to track.

10.7. Assessing epistemological objections to illusionism

Consider again the anti-illusionist epistemological arguments. Our situation in the epistemological domain does not exactly mirror our situation in the ethical domain. Serious epistemological revisions appear as soon as we combine illusionism with non-phenomenalist views of justification. No non-phenomenalist view depicts entities playing the evidential role played by phenomenal experiences in our intuitive picture. If Meta-Cartesianism is correct, this is unsurprising: we essentially represent phenomenal experiences *as entities playing a peculiar epistemological role*. A non-phenomenalist view of justification positing mental states playing the exact same epistemological role must be a phenomenalist view in disguise. Any epistemological view, combined with illusionism, brings serious revisions.

Does this mean that anti-illusionist arguments appealing to epistemological data succeed? No. As there are plausible non-phenomenalist views of justification, illusionists can maintain that things other than phenomenal experiences ground justification. They can reject exclusive phenomenalism and avoid epistemological nihilism. Non-phenomenalist epistemological views can account for alleged manifest truths about justification, and

explain, say, that we are immediately justified to believe various things about our perceptible environment.

What illusionists have to give up is introspective obviousness: we do not have mental states which, like phenomenal experiences on the intuitive picture, are introspectively obvious—self-revealing and infallibly introspectively presented. There is no mental domain "where appearances are, not just a *promise* of truth, but a *guarantee* of truth."[43]

However, while the idea that some of our mental states are introspectively obvious is intuitive, I contend that, all things considered, it should not be part of our basic epistemological data—the data that any view of justification must account for.

First, this idea has long been criticized independently of illusionism—by coherentists, externalist foundationalists, or weak internalist foundationalists. Its acceptance should not be the touchstone of our epistemology. Most contemporary epistemologists would agree.

Second, even granting that it is *manifest* that some mental states are introspectively obvious, I argued in Chapter 7, when discussing the Moorean argument against illusionism, that we have good science-based arguments against this claim, and that it is ultimately implausible that this claim is so manifest that it still counts as data once these arguments are considered. An anti-illusionist argument appealing to our manifest possession of substantive infallible justification about some of our mental states would roughly amount to a Moorean argument for something like acquaintance with these mental states, which I addressed when discussing the Moorean argument against illusionism.

Another anti-illusionist epistemological argument is that, since illusionists deny infallible justification, they must accept that they cannot conclusively rule out absolute *nihilism*—the view that *nothing* exists at all. However, it seems "unreasonable to wonder whether absolute nihilism is true."[44]

While illusionism naturally leads to denying infallible justification, it is in good company, since many other epistemological views have the same consequence. Fallibilism regarding justification is popular.[45] Moreover, while it is indeed unreasonable to wonder whether absolute nihilism is true, this does not imply that we have infallible justification to believe that some specific entity exists. Other considerations account for this.

[43] Merlo (2020, 129).
[44] C. D. Brown (2022, 20).
[45] Hetherington (2024).

First, coherentist considerations: absolute nihilism contradicts virtually all of our beliefs about existent things, and it is hard to think of an *internally coherent* view that is *less coherent* with our prior beliefs. On coherentist grounds, it is unreasonable to wonder whether anything exists. Second, considerations pertaining to the possibility of a justificatory path. Even if all our justifications are defeasible, we should only view them as defeated when we have reasons to do so. These reasons would consist of arguments, like debunking arguments, presupposing the existence of *some* things. If so, absolute nihilism is impossible to justify. There is no possible justificatory route leading to it, since argumentative steps of any such route presuppose its falsity. Therefore, it is unreasonable to seriously wonder whether absolute nihilism is true.

Anti-illusionist arguments based on epistemological data fail. Illusionism leads to some epistemological revisions, but it does not contradict what should ultimately count as epistemological data.

* * *

As illusionists, we must abandon our intuitive picture of justification. This leads to revisions not only about the grounds of justification, but also about the type of justification we really have. The nature of these revisions depends on the non-phenomenalist view of justification we accept. I presented three non-phenomenalist views of justification, all of them revisionary to some extent, but which still account for our ultima facie epistemological data. Epistemological objections to illusionism fail.

Coda

Consciousness, Illusion, and Reality

We reach the end of our journey. What have we discovered?

Phenomenal consciousness appears mysterious—obvious, irreducible, undetectable—and highly significant (Chapter 1). However, this form of consciousness does not really exist, but only seems to exist. Illusionism is true (Chapters 2 and 3).

Why does it seem to exist? Currently available answers to this question fail. They do not explain why phenomenal consciousness seems obvious (Chapter 4). To explain this, the evidential approach is needed. We must understand the tight link between phenomenal cognition and evidential cognition. Meta-Cartesianism is a version of this approach (Chapter 5). It explains why phenomenal consciousness seems to exist, with all its peculiar features—obviousness, irreducibility, undetectability, and significance (Chapter 6). Besides, objections to illusionism can be answered (Chapter 7).

Illusionism has implications in various domains. First, consciousness science. Illusionism does not imply that consciousness science is moot, but it does not leave it untouched either. It has non-trivial implications and suggests that certain questions and approaches should be reconsidered (Chapter 8). Illusionism also has ethical implications for prudential value and moral status. It does not entail nihilism about those, but it implies a revision of our intuitive ethical picture (Chapter 9). Finally, it has epistemological implications. Again, without implying that justifications do not exist, it requires revising our intuitive epistemological picture (Chapter 10).

* * *

We are physical creatures. We evolved through natural selection. In this process, we acquired more and more sophisticated cognitive capacities. Among them was the ability to represent aspects of our own cognition and notably what *appears* to us: what counts as evidence for us, and particularly as *direct evidence*, i.e., evidence we have independently of our beliefs.

House of Mirrors. François Kammerer, Oxford University Press. © Oxford University Press 2026.
DOI: 10.1093/9780197840283.003.0012

We gained this ability by acquiring convenient misrepresentations: we came to represent that things appear to us through the medium of peculiar states, phenomenal experiences, endowed with a unique epistemological status making them introspectively obvious. However, none of our mental states really has this status. There are no phenomenal experiences.

This story, told in the third person, seems reasonable. Should we not expect that evolving cognitive systems would come to represent themselves in mischaracterizing ways? A complete *absence* of mischaracterization would be surprising. The idea that our representations of our own minds imply mischaracterizations is banal: from the third-person perspective, it is akin to the banal idea that evolving cognitive systems mischaracterize their perceptible environment in useful ways—e.g., represent primitive sensible qualities that do not really exist.

Our introspective mischaracterizations uniquely concern the evidence at our disposal. They are about the tool we use when, engaged in inquiry, we represent our epistemic situation: when we represent what appears to us and what we should believe. This generates unique intuitive limitations when considering that phenomenal consciousness fails to exist and only seems to exist. The story told above, reasonable and banal from the third-person perspective, appears absurd about ourselves. This makes the illusion of phenomenal consciousness uniquely hard to apprehend as such.

Our most intimate apprehensions of what happens in us are riddled with a profound error. Our deepest acts of introspection are mischaracterizing. We should not forget how surprising and even scandalous this conclusion is. Since the beginning of philosophy, East and West, many have claimed that the world, in some sense, is an illusion. Some concluded that, to avoid falling prey to this illusion, we should ignore the world and focus on our minds. However, our minds, too, and therefore, illusions as we conceive of them, are illusions. Experience, the medium in which appearances are supposedly given, is not as we represent it: it does not involve these phenomenal properties we introspectively represent. Retreating to the intimacy of our experiences does not make us escape mirages. Trying to escape Plato's Cave, we entered Descartes's House of Mirrors; there, too, we were faced with misleading shadows, and their dizzying reflections made the House uniquely hard to escape.

The philosopher who has ever had the nerve to defend illusionism publicly often encounters, from colleagues or strangers, half-fascinated and half-disgusted incredulous stares, frequently followed by a question: Do you

really believe *this*? Here is an honest answer. The view that *I* don't have any phenomenal experiences seems somehow absurd *to me too*, when I contemplate it intuitively. The view makes sense from the third-person perspective; from this perspective, I even think it is the one that makes the most sense. It predicts that it should seem absurd from the first-person perspective. But sometimes, things that seem absurd can be true, and I think that we are probably in one of these cases. In this sense, I really believe this.

We *can* understand the introspective illusion of phenomenal consciousness. We can bootstrap ourselves and acknowledge that we do not really have the phenomenal experiences we introspect, even if it seems absurd. Doing so, we acknowledge that the illusion of phenomenal consciousness is not a chaotic hallucination disconnected from reality. The same way the outside world really contains non-primitive and non-qualitative properties we mischaracterize as primitive qualities, our minds really bear non-phenomenal properties we mischaracterize as phenomenal. We really have minds, and we really are conscious in various modest senses. Simply, we do not have these unique and mysterious phenomenal experiences that so strongly seem real to us.

Because phenomenal consciousness seems so central to what we take ourselves to be and so significant, understanding its non-existence has implications. These implications should not be caricatured, and simplistic slogans, from the radical "everything must go" to the quietist "business as usual," are to be avoided. We must understand, patiently and cautiously, how admitting the illusory nature of the phenomenal suggests (but rarely forces) subtle revisions of our ways of studying the mind, thinking about value and moral status, and conceiving of justification. In this book, I presented what I hope are helpful analogies.

The science of consciousness is like a *science of sirens*. Its real objects are, on the one hand, ordinary but mischaracterized entities: quasi-phenomenal consciousness, or manatees. On the other hand, they are the mischaracterizing representations of these entities: illusions of phenomenal consciousness, or illusory sightings of sirens. The same way marine biologists should not let their study of manatees be shaped by their understanding of mythical sirens, consciousness scientists friendly to illusionism should not let their idea of *phenomenal consciousness* shape their study of quasi-phenomenal consciousness.

In ethics, illusionists are like *newly convinced atheists*. Recent atheists should not become nihilists simply because they admit that what they thought of as the basis of ethics—God—does not exist. Their ethical thinking

should not be shaped by their lingering theistic tendencies. The same is true for illusionists, who should guard themselves against lingering sentientist habits. Humans and other creatures are able to care about various things and can engage in various good activities. This is enough to make their lives go well or badly and make them deserve moral consideration, even if they lack phenomenal experiences.

Finally, in epistemology, the phenomenal conception of justification was a useful tool, with which we conveniently made sense of our epistemic situations. It has met its limits. Like the old skin of a reptile in shed, it must be discarded in favor of non-phenomenalist views of justification, which can better structure our way of thinking about our epistemic situations.

* * *

Illusionism concerns what does *not* exist. Its only positive commitment regards the illusion of phenomenal consciousness. However, I claimed that this illusion is the deployment of some physical cognitive mechanism. Moreover, my argument from illusionism presupposes the existence of the physical. Being an illusionist without recognizing that the physical exists makes no sense. Illusionism does not imply physicalism, but since physicalism is independently attractive, and since illusionism removes one of the biggest threats to physicalism, it seems natural for illusionists to be physicalists.

Earlier, I primarily understood "physical" as meaning "exhaustively describable in terms of structure and dynamics." This yields a sense of physicalism—structural physicalism—which makes it equivalent to *ontic structural realism* in the metaphysics of science,[1] the view that, roughly, structure is all there is to reality. For an ontic structural realist, there are, at the fundamental level, no individual objects with intrinsic and categorical properties. To the extent that there are objects, they are just nodes in a structure.

A frequent objection to ontic structural realism is that the claim that only structures exist is unintelligible. The existence of structures conceptually requires underlying objects forming these structures. These objects, one might think, must be a certain way in themselves; they must have certain fundamental, intrinsic, and categorical features, which are not, ultimately, structurally constituted: quiddities. One could challenge the illusionist, who is naturally led to structural physicalism, by contending that this view is unintelligible for the same reason.

[1] Ladyman (1998); Esfeld (2009); French and Ladyman (2010).

In Chapter 2, I discussed the concept EXIST, required to grasp the opposition between phenomenal realism and illusionism. I claimed that EXIST is a primitive concept we all grasp. This objection to structural physicalism I just presented can be interpreted as relying on the view that our primitive concept of existence requires existent things to be ultimately grounded in the instantiation of properties that are fundamental, intrinsic, and categorical.

As I said, this objection does not target illusionism per se, only structural physicalism, which illusionists can coherently deny. However, one might add that fundamental intrinsic categorical properties *must be* phenomenal. After all, phenomenal properties might be the best (if not the only) example of fundamental intrinsic categorical properties we can think of. If such properties have to be phenomenal, illusionism is committed to denying their existence. Taking the earlier objection for granted, this would imply that illusionism is unintelligible.

The idea that our concept of existence is such that the non-existence of the phenomenal is unintelligible has been historically considered. Take Raymond Ruyer, a highly original (and almost forgotten) French metaphysician of the mid-20th century. His first book defended a version of ontic structural realism, on which

> There is only one kind of reality: geometrico-mecanical reality, form, structure. The whole diversity of the world stems exclusively from the diversity of forms.[2]

Ruyer then abandoned this view and came to give a central role to subjective entities as the realizer of these structures, accepting a Russellian conception. Arguably, he did so because he saw the need to admit the existence of phenomenal consciousness as a primary constraint stemming from our concept of existence or reality. Discussing the ontological relations between a sensation of light and its cerebral basis, he wrote:

> We have the intuition of the absolute reality of the sensation [. . .]. [The] sensation of light is incontestably a part of reality, *or the word reality no longer means anything*.[3]

[2] Ruyer (1930, 2, my translation).
[3] Ruyer ([1937] 2020, 34, my translation, my emphasis).

Ruyer is an example among others. American philosopher D.C. Williams expressed a comparable idea when criticizing the view, which he associated with behaviorism, that consciousness does not exist:

> Not only does the conscious content exist; it is for me *the very type of existence*.[4]

British idealist F.H. Bradley thought that the way we conceive of the *real* requires real things to have certain links to phenomenal experience:

> We perceive, on reflection, that to be real, or even barely to exist, must be to fall within sentience. Sentient experience, in short, is reality, and what is not this is not real [. . .]. Anything, in no sense felt or perceived, becomes to me quite unmeaning. [. . .] [It] seems, in my mind, to be a mere word and a failure, or else an attempt at self-contradiction. It is a vicious abstraction whose existence is meaningless nonsense, and is therefore not possible.[5]

Of course, illusionists can concede that EXIST requires existent things to be grounded in the instantiation of fundamental, intrinsic, and categorical features, and renounce structural physicalism, while remaining illusionists, provided they deny that fundamental, intrinsic, and categorical properties must be phenomenal or grounded in the phenomenal. They could do so and remain physicalists in the weaker, purely negative sense of the term.

Suppose illusionists do not allow themselves this answer. Suppose they concede that fundamental, intrinsic, and categorical properties must be phenomenal, or grounded in the phenomenal.[6] In this case, the choice ends up being between illusionism coupled with something like structural physicalism, and phenomenal realism. How could illusionists then react to the objection that EXIST requires existent things to be grounded in something fundamental, intrinsic, and categorical—and therefore in something phenomenal?

They could, first, deny this claim about EXIST. Grant, however, that the claim has intuitive pull. Another possibility would be to admit that we should revise our conception of existence: not merely change our view of *what*

[4] Williams (1934, 465, my emphasis).
[5] Bradley (1893, XIV). The citation is taken from Builes (2024).
[6] See the arguments in this direction given in Builes (2024).

exists, but also our view of *what it means to exist*. Illusionists could try to disentangle strands knit together in our primitive concept EXIST—say, having causal powers on the one hand, having some intrinsic categorical nature on the other. They could state that the *physical* exists only in the first sense; in the second sense, nothing at all exists.

Read liberally, the idea that reflecting on consciousness leads to revising or correcting our conception of existence is not new. Husserl thought that Descartes's cogito was on the right track, but that Descartes went astray when he conceived of thinking as inhering in an existing thing—a *res*. For Husserl, understanding consciousness requires grasping it, not as a thing among things, but as a condition of possibility of things.[7] Neuroscientist Nick Humphrey, whose work is friendly to illusionism, refused the illusionist label, for rhetorical reasons discussed previously, but his rhetorical argument can also be interpreted in a substantive and metaphysical manner. Humphrey claimed to defend an "amphibian theory" of consciousness in the sense of Leibniz, for whom imaginary numbers were

> an invention . . . a marvel of analysis, a prodigy of the world of ideas, almost an amphibian between Being and Non-Being.[8]

If phenomenal consciousness is amphibian, if it is between existence and non-existence, properly understanding it requires revising our concept of existence and grasping that phenomenal consciousness exists in some sense but not in others.

I do not endorse any of these views. Still, I find the idea that our conception of what it is to exist could have to be revised when we adopt something like illusionism stimulating. Our way of thinking of value, moral status, and epistemic justification changes when we cease to understand all these features as grounded in the phenomenal. It is not implausible that our way of thinking of existence itself also changes once the phenomenal ceases to constitute, for us, a paradigmatic part of reality. A true philosophical journey rarely leaves one's starting point unchanged.

[7] Husserl ([1931] 1960, §§9–11).
[8] Humphrey (2019, 106).

References

Adams, Robert Merrihew. 1987. *The Virtue of Faith and Other Essays in Philosophical Theology*. Oxford University Press.

Albantakis, Larissa, Leonardo Barbosa, Graham Findlay, et al. 2023. "Integrated Information Theory (IIT) 4.0: Formulating the Properties of Phenomenal Existence in Physical Terms." *PLOS Computational Biology* 19 (10): e1011465. https://doi.org/10.1371/journal.pcbi.1011465.

Albrecht, Gary L., and Patrick J. Devlieger. 1999. "The Disability Paradox: High Quality of Life against All Odds." *Social Science & Medicine* 48 (8): 977–88. https://doi.org/10.1016/S0277-9536(98)00411-0.

Allen, Colin. 2013. "Fish Cognition and Consciousness." *Journal of Agricultural and Environmental Ethics* 26 (1): 25–39. https://doi.org/10.1007/s10806-011-9364-9.

Allen, Colin, and Mark Bekoff. 1997. *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. MIT Press.

Alston, William. 1971. "Varieties of Privileged Access." *American Philosophical Quarterly* 8 (3): 223–41.

Alston, William. 2005. *Beyond Justification: Dimensions of Epistemic Evaluation*. Cornell University Press.

Alter, Torin. 2016. "The Structure and Dynamics Argument against Materialism." *Noûs* 50 (4): 794–815. https://doi.org/10.1111/nous.12134.

Andrews, Kristin. 2024. "'All Animals Are Conscious': Shifting the Null Hypothesis in Consciousness Science." *Mind & Language* 39 (January): 415–33. https://doi.org/10.1111/mila.12498.

Andrews, Kristin, Jonathan Birch, Jeff Sebo, and Toni Sims. 2024. *Background to the New York Declaration on Animal Consciousness*. nydeclaration.com.

Antony, Michael. 2006. "Vagueness and the Metaphysics of Consciousness." *Philosophical Studies* 128 (3): 515–38. https://doi.org/10.1007/s11098-004-7488-8.

Appleby, M.C., and P. Sandøe. 2002. "Philosophical Debate on the Nature of Well-Being: Implications for Animal Welfare." *Animal Welfare* 11 (3): 283–94. https://doi.org/10.1017/S0962728600024866.

Aristotle. 2016. *Metaphysics*. Translated by C.D.C. Reeve. Hackett Publishing Company.

Armstrong, David. 1961. *Perception and the Physical World*. Routledge & Kegan.

Armstrong, David. 1968. "The Headless Woman Illusion and the Defence of Materialism." *Analysis* 29 (2): 48–49. https://doi.org/10.1093/analys/29.2.48.

Armstrong, David. 1980. *The Nature of Mind and Other Essays*. Cornell University Press.

Artiga, Marc. 2021. "Beyond Black Dots and Nutritious Things: A Solution to the Indeterminacy Problem." *Mind & Language* 36 (3): 471–90. https://doi.org/10.1111/mila.12284.

Augustine. 1990. *The Trinity*. Edited by John E. Rotelle. Translated by Edmund Hill. New City Press.

Augustine. 1993. *Against the Academicians (Contra Academicos)*. 6th printing. Translated by Mary Patricia Garvey. Mediaeval Philosophical Texts in Translation 2. Marquette University Press.

Aydede, Murat. 2019. "Pain." In *The Stanford Encyclopedia of Philosophy*, Spring 2019, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/pain/.

Aydede, Murat, and Güven Güzeldere. 2005. "Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness." *Noûs* 39 (2): 197–255. https://doi.org/10.1111/j.0029-4624.2005.00500.x.

Baars, Bernard. 1988. *A Cognitive Theory of Consciousness*. Cambridge University Press.

Baddeley, Alan. 1999. "Memory." In *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press.

Bahrami, Bahador, Karsten Olsen, Peter E. Latham, Andreas Roepstorff, Geraint Rees, and Chris D. Frith. 2010. "Optimally Interacting Minds." *Science* 329 (5995): 1081–85. https://doi.org/10.1126/science.1185718.

Balog, Katalin. 2009. "Phenomenal Concepts." In *Oxford Handbook in the Philosophy of Mind*, edited by Brian McLaughlin, Ansgar Beckermann, and Sven Walter. Oxford University Press.

Balog, Katalin. 2012a. "Acquaintance and the Mind-Body Problem." In *New Perspectives on Type Identity: The Mental and the Physical*, edited by Christopher Hill and Simone Gozzano. Cambridge University Press.

Balog, Katalin. 2012b. "In Defense of the Phenomenal Concept Strategy." *Philosophy and Phenomenological Research* 84 (1): 1–23. https://doi.org/10.1111/j.1933-1592.2011.00541.x.

Balog, Katalin. 2016. "Illusionism's Discontent." *Journal of Consciousness Studies* 23 (11–12): 40–51.

Balog, Katalin. 2023. "The Abolition of Phenomena: A Voyage among Zombies." *Klesis* 55: 1–19.

Bang, Dan, Riccardo Fusaroli, Kristian Tylén, et al. 2014. "Does Interaction Matter? Testing Whether a Confidence Heuristic Can Replace Interaction in Collective Decision-Making." *Consciousness and Cognition* 26 (May): 13–23. https://doi.org/10.1016/j.concog.2014.02.002.

Banse, Karl. 1990. "Mermaids—Their Biology, Culture and Demise." *Limnology and Oceanography* 35 (1): 148–53. https://doi.org/10.4319/lo.1990.35.1.0148.

Barrett, L.F., and Moshe Bar. 2009. "See It with Feeling: Affective Predictions during Object Perception." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1521): 1325–34. https://doi.org/10.1098/rstb.2008.0312.

Bayertz, Kurt, Myriam Gerhard, and Walter Jaeschke. 2007. *Der Materialismus-Streit*. Weltanschauung, Philosophie und Naturwissenschaft im 19. Jahrhundert, Band 1. F. Meiner.

Bayle, Pierre. 1699. *Pensées Diverses Écrites à Un Docteur de Sorbonne à l'occasion de La Comète Qui Parut Au Mois de Décembre 1680*. 3rd ed. Reinier Leers.

Bayle, Pierre. [1702] 1735. *The Dictionary Historical and Critical*. 2nd ed. Translated by Pierre des Maizeaux. Vol 2. Knapton et al.

Bayne, Tim, Joel Frohlich, Rhodri Cusack, Julia Moser, and Lorina Naci. 2023. "Consciousness in the Cradle: On the Emergence of Infant Experience." *Trends in Cognitive Sciences* 27 (12): 1135–49. https://doi.org/10.1016/j.tics.2023.08.018.

Bayne, Tim, Jakob Hohwy, and Adrian M. Owen. 2016. "Are There Levels of Consciousness?" *Trends in Cognitive Sciences* 20 (6): 405–13. https://doi.org/10.1016/j.tics.2016.03.009.

Bayne, Tim, and Michelle Montague, eds. 2011. *Cognitive Phenomenology*. Oxford University Press.

Bayne, Tim, Anil K. Seth, Marcello Massimini, et al. 2024. "Tests for Consciousness in Humans and Beyond." *Trends in Cognitive Sciences* 28 (5): 454–66. https://doi.org/10.1016/j.tics.2024.01.010.

Bayne, Tim, and Nicholas Shea. 2020. "Consciousness, Concepts, and Natural Kinds." *Philosophical Topics* 48 (1): 65–84. https://doi.org/10.5840/philtopics20204814.

BBC News. 2016. "Indonesia Police Confiscate Sex Toy Mistaken for 'Angel.'" Asia. *BBC News*, May 3. https://www.bbc.com/news/world-asia-36189614.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March 3, 610–23. https://doi.org/10.1145/3442188.3445922.

Bentham, Jeremy. [1789] 1961. *An Introduction to the Principles of Morals and Legislation*. Doubleday.

Beran, Michael. 2019. "Animal Metacognition: A Decade of Progress, Problems, and the Development of New Prospects." *Animal Behavior and Cognition* 6 (4): 223–29. https://doi.org/10.26451/abc.06.04.01.2019.

Berger, Jacob. 2020. "Perceptual Consciousness Plays No Epistemic Role." *Philosophical Issues* 30 (1): 7–23. https://doi.org/10.1111/phis.12168.

Berger, Jacob, Bence Nanay, and Jake Quilty-Dunn. 2018. "Unconscious Perceptual Justification." *Inquiry: An Interdisciplinary Journal of Philosophy* 61 (5): 569–89. https://doi.org/10.1080/0020174x.2018.1432413.

Bergson, Henri. [1922] 1999. *Duration and Simultaneity: Bergson and the Einsteinian Universe*. Edited by Robin Durie. Clinamen Press.

Bermond, B. 2001. "A Neuropsychological and Evolutionary Approach to Animal Consciousness and Animal Suffering." *Animal Welfare* 10 (S1): S47–62. https://doi.org/10.1017/S0962728600023514.

Billon, Alexandre. 2015. "Why Are We Certain That We Exist?" *Philosophy and Phenomenological Research* 91 (3): 723–59. https://doi.org/10.1111/phpr.12113.

Birch, Jonathan. 2022a. "Materialism and the Moral Status of Animals." *Philosophical Quarterly* 72 (4): 795–815. https://doi.org/10.1093/pq/pqab072.

Birch, Jonathan. 2022b. "The Search for Invertebrate Consciousness." *Noûs* 56 (1): 133–53. https://doi.org/10.1111/nous.12351.

Birch, Jonathan. 2024. *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*. Oxford Scholarship Online. Oxford University Press. https://doi.org/10.1093/9780191966729.001.0001.

Birch, Jonathan, Donald M. Broom, Heather Browning, et al. 2022. "How Should We Study Animal Consciousness Scientifically?" *Journal of Consciousness Studies* 29 (3): 8–28. https://doi.org/10.53765/20512201.29.3.008.

Birch, Jonathan, Alexandra K. Schnell, and Nicola S. Clayton. 2020. "Dimensions of Animal Consciousness." *Trends in Cognitive Sciences* 24 (10): 789–801. https://doi.org/10.1016/j.tics.2020.07.007.

Block, Ned. 1978. "Troubles with Functionalism." *Minnesota Studies in the Philosophy of Science* 9 (261–325): 261–325.

Block, Ned. 1986. "Advertisement for a Semantics for Psychology." *Midwest Studies in Philosophy* 10 (1): 615–78.

Block, Ned. 1995. "On a Confusion about a Function of Consciousness." *Brain and Behavioral Sciences* 18: 227–47. https://doi.org/10.1017/S0140525X00038188.

Block, Ned. 2002. "The Harder Problem of Consciousness." *Journal of Philosophy* 99 (8): 391–425. https://doi.org/10.2307/3655621.

Block, Ned. 2007. "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience." *Behavioral and Brain Sciences* 30: 481–548. https://doi.org/10.1017/S0140525X07002786.

Block, Ned. 2009. "Comparing the Major Theories of Consciousness." In *The Cognitive Neurosciences*, 4th ed., edited by Michael S. Gazzaniga. MIT Press. https://doi.org/10.7551/mitpress/8029.003.0099.

Block, Ned. 2019. "What is wrong with the o-report paradigm and how to fix it." *Trends in Cognitive Sciences* 23 (12): 1003–1013. https://doi.org/10.1016/j.tics.2019.10.001.

Bloom, Paul. 2018. *Against Empathy: The Case for Rational Compassion*. Vintage.

Blough, Donald S. 2001. "The Perception of Similarity." In *Avian Visual Cognition*, edited by Robert G. Cook. Comparative Cognition Press. pigeon.psy.tufts.edu/avc/dblough.

Bois-Reymond, Emil Heinrich du. 1872. *Über Die Grenzen Des Naturerkennens*. Veit.

Boly, Melanie, Anil K. Seth, Melanie Wilke, et al. 2013. "Consciousness in Humans and Non-Human Animals: Recent Advances and Future Directions." *Frontiers in Psychology* 4. https://doi.org/10.3389/fpsyg.2013.00625.

BonJour, Laurence. 2000. "Toward a Defense of Empirical Foundationalism." In *Resurrecting Old-Fashioned Foundationalism*, edited by Michael DePaul. Rowman and Littlefield.
Boult, Cameron. 2017. "Epistemic Normativity and the Justification-Excuse Distinction." *Synthese* 194 (10): 4065–81. https://doi.org/10.1007/s11229-016-1127-8.
Boyle, Alexandria. 2022. "Episodic Memory in Animals: Optimism, Kind Scepticism and Pluralism." In *Current Controversies in Philosophy of Memory*. Routledge.
Bradford, Gwen. 2015. "Perfectionism." In *The Routledge Handbook of Philosophy of Well-Being*, edited by Guy Fletcher. Routledge.
Bradford, Gwen. 2021. "Perfectionist Bads." *The Philosophical Quarterly* 71 (3): 586–604. https://doi.org/10.1093/pq/pqaa055.
Bradford, Gwen. 2023. "Consciousness and Welfare Subjectivity." *Noûs* 57 (4): 905–21. https://doi.org/10.1111/nous.12434.
Bradley, Francis Herbert. 1893. *Appearance and Reality*. Swan Sonnenschein.
Braithwaite, Victoria. 2010. *Do Fish Feel Pain?* Oxford University Press.
Brentano, Franz. 1952. *Grundlegung Und Aufbau Der Ethik, Nach Den Vorlesungen Über Praktische Philosophie*. Meiner.
Broad, Charlie Dunbar. 1925. *The Mind and Its Place in Nature*. Routledge & Kegan.
Brogaard, Berit. 2018. "Phenomenal Dogmatism, Seeming Evidentialism and Inferential Justification." In *Believing in Accordance with the Evidence*, edited by Kevin McCain, vol. 398. Synthese Library. Springer International Publishing. https://doi.org/10.1007/978-3-319-95993-1_5.
Bronfman, Zohar Z., Simona Ginsburg, and Eva Jablonka. 2016. "The Transition to Minimal Consciousness through the Evolution of Associative Learning." *Frontiers in Psychology* 7 (December). https://doi.org/10.3389/fpsyg.2016.01954.
Broussais, François-Joseph-Victor. 1828. *De l'irritation et de La Folie*. Delaunay.
Brown, Christopher Devlin. 2022. "Why Illusionism about Consciousness Is Unbelievable." *Ratio* 35 (1): 16–24. https://doi.org/10.1111/rati.12321.
Brown, Christopher Devlin, and David Papineau. 2024. "Illusionism and A Posteriori Physicalism: No Fact of the Matter." *Journal of Consciousness Studies* 31 (7–8): 7–27. https://doi.org/10.53765/20512201.31.7.007.
Brown, Richard. 2010. "Deprioritizing the A Priori Arguments against Physicalism." *Journal of Consciousness Studies* 17 (3–4): 47–69.
Brown, Richard, Hakwan Lau, and Joseph E. LeDoux. 2019. "Understanding the Higher-Order Approach to Consciousness." *Trends in Cognitive Sciences* 23 (9): 754–68. https://doi.org/10.1016/j.tics.2019.06.009.
Browning, Heather, and Walter Veit. 2020. "The Measurement Problem of Consciousness." *Philosophical Topics* 48 (1): 85–108.
Browning, Heather, and Walter Veit. 2023. "Studying Introspection in Animals and AIs." *Journal of Consciousness Studies* 30 (9): 63–74. https://doi.org/10.53765/20512201.30.9.063.
Builes, David. 2024. "Modal Idealism." In *Oxford Studies in Philosophy of Mind*, edited by Uriah Kriegel. Oxford University Press. https://doi.org/10.1093/9780198924159.003.0002.
Burnyeat, Myles Fredric. 1982. "Idealism and Greek Philosophy: What Descartes Saw and Berkeley Missed." *The Philosophical Review* 91 (1): 3–40.
Butlin, Patrick, Robert Long, Eric Elmoznino, et al. 2023. *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*. Version 3. https://doi.org/10.48550/ARXIV.2308.08708.
Cabanac, Michel, Arnaud J. Cabanac, and André Parent. 2009. "The Emergence of Consciousness in Phylogeny." *Behavioural Brain Research* 198 (2): 267–72. https://doi.org/10.1016/j.bbr.2008.11.028.
Cabanis, Pierre Jean Georges. 1805. *Rapports Du Physique et Du Moral de l'homme*. 2nd ed. Crapart, Caille et Ravier.
Cardaillac, Séverin de. 1830. *Etudes Élémentaires de Philosophie*. Vol. 1. Firmin Didot.

Carey, Susan, and Elisabeth Spelke. 1996. "Science and Core Knowledge." *Philosophy of Science* 63 (4): 515–33.

Carruthers, Peter. 1992. *The Animals Issue: Moral Theory in Practice*. Cambridge University Press.

Carruthers, Peter. 1999. "Sympathy and Subjectivity." *Australasian Journal of Philosophy* 77 (4): 465–82. https://doi.org/10.1080/00048409912349231.

Carruthers, Peter. 2000. *Phenomenal Consciousness*. Cambridge University Press.

Carruthers, Peter. 2004. "Phenomenal Concepts and Higher-Order Experiences." *Philosophy and Phenomenological Research* 68 (2): 316–36. https://doi.org/10.1111/j.1933-1592.2004.tb00343.x.

Carruthers, Peter. 2005a. "Conscious Experience versus Conscious Thought." In *Consciousness: Essays from a Higher-Order Perspective*. Oxford University Press.

Carruthers, Peter. 2005b. *Consciousness: Essays from a Higher-Order Perspective*. Clarendon Press.

Carruthers, Peter. 2018. "Valence and Value." *Philosophy and Phenomenological Research* 97 (3): 658–80. https://doi.org/10.1111/phpr.12395.

Carruthers, Peter. 2019. *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford University Press.

Carruthers, Peter. 2020. "Stop Caring about Consciousness." *Philosophical Topics* 48 (1): 1–20. https://doi.org/10.5840/philtopics20204811.

Carruthers, Peter. 2023. "On Valence: Imperative or Representation of Value?" *The British Journal for the Philosophy of Science* 74 (3): 533–53. https://doi.org/10.1086/714985.

Carruthers, Peter, and David M. Williams. 2019. "Comparative Metacognition." *Animal Behavior and Cognition* 6 (4): 278–88. https://doi.org/10.26451/abc.06.04.08.2019.

Carter, Olivia, Jakob Hohwy, Jeroen Van Boxtel, et al. 2018. "Conscious Machines: Defining Questions." *Science* 359 (6374): 400. https://doi.org/10.1126/science.aar4163.

Caston, V. 2002. "Aristotle on Consciousness." *Mind* 111 (444): 751–815. https://doi.org/10.1093/mind/111.444.751.

Caston, Victor. 2015. "Perception in Ancient Greek Philosophy." *The Oxford Handbook of Philosophy of Perception*, edited by Mohan Matthen, 29–50. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199600472.013.036.

Cermeño-Aínsa, Sergio. 2024. "Yogācāra Buddhism and the Illusion of Phenomenal Consciousness." *Asian Philosophy* 35 (2): 161–80. https://doi.org/10.1080/09552367.2024.2396746.

Chalmers, David. 1995. "Facing up to the Problem of Consciousness." *Journal of Consciousness Studies* 2 (3): 200–219.

Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, David. 2000. "What Is a Neural Correlate of Consciousness?" In *Neural Correlates of Consciousness*, edited by Thomas Metzinger. MIT Press.

Chalmers, David. 2002a. "Consciousness and Its Place in Nature." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David Chalmers. Oxford University Press.

Chalmers, David. 2002b. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility*, edited by Tamar Szabo Gendler and John Hawthorne. Oxford University Press.

Chalmers, David. 2003. "The Content and Epistemology of Phenomenal Belief." In *Consciousness: New Philosophical Perspectives*, edited by Quentin Smith and Aleksandar Jokic. Oxford University Press.

Chalmers, David. 2004. "The Representational Character of Experience." In *The Future for Philosophy*, edited by Brian Leiter. Oxford University Press. https://doi.org/10.1093/oso/9780199247288.003.0007.

Chalmers, David. 2006. "Perception and the Fall from Eden." In *Perceptual Experience*, edited by Tamar Szabo and John Hawthorne. Oxford University Press.

Chalmers, David. 2015. "Panpsychism and Panprotopsychism." *Consciousness in the Physical World: Perspectives on Russellian Monism* 246 (2003): 102–54.
Chalmers, David. 2016. "The Combination Problem for Panpsychism." In *Panpsychism*, edited by Godehard Brüntrup and Ludwig Jaskolla. Oxford University Press.
Chalmers, David. 2018. "The Meta-Problem of Consciousness." *Journal of Consciousness Studies* 25 (9–10): 6–61.
Chalmers, David. 2020a. "Debunking Arguments for Illusionism about Consciousness." *Journal of Consciousness Studies* 27 (5–6): 258–81.
Chalmers, David. 2020b. "How Can We Solve the Meta-Problem of Consciousness." *Journal of Consciousness Studies* 27 (5–6): 201–26.
Chalmers, David. 2020c. "Is the Hard Problem of Consciousness Universal?" *Journal of Consciousness Studies* 27 (5–6): 227–57.
Chalmers, David. 2022. *Reality+: Virtual Worlds and the Problems of Philosophy*. W.W. Norton & Company.
Chalmers, David, and David Bourget. 2020. "PhilPapers Survey 2020." https://survey2020.philpeople.org/.
Chalmers, David, and Frank Jackson. 2001. "Conceptual Analysis and Reductive Explanation." *Philosophical Review* 110: 315–61. https://doi.org/10.2307/2693648.
Chalmers, David, and Kelvin McQueen. 2022. "Consciousness and the Collapse of the Wave Function." In *Consciousness and Quantum Mechanics*, edited by Shan Gao. Oxford University Press.
Chaturvedi, Amit. 2023. "Is the Mind a Magic Trick? Illusionism about Consciousness in the 'Consciousness-Only' Theory of Vasubandhu and Sthiramati." *Ergo* 10 (52): 1495–534. https://doi.org/10.3998/ergo.5189.
Chisholm, Roderick. 1957. *Perceiving: A Philosophical Study*. Cornell University Press.
Chrisley, Ron, and Aaron Sloman. 2016. "Functionalism, Revisionism, and Qualia." *APA Newsletter on Philosophy and Computers* 16 (1): 2–13.
Chudnoff, Elijah. 2012. "Presentational Phenomenology." In *Consciousness and Subjectivity*, edited by Sofia Miguens and Gerhard Preyer. Ontos Verlag.
Chudnoff, Elijah. 2014. *Intuition*. Oxford University Press.
Churchland, Paul M. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78 (2): 67–90. https://doi.org/10.2307/2025900.
Churchland, Paul M. 1985. "Reduction, Qualia, and the Direct Introspection of Brain States." *Journal of Philosophy* 82 (1): 8–28. https://doi.org/10.2307/2026509.
Clark, Andy. 2000. "A Case Where Access Implies Qualia?" *Analysis* 60: 30–38. https://doi.org/10.1093/analys/60.1.30.
Clark, Andy, and David Chalmers. 1998. "The Extended Mind." *Analysis* 58 (1): 7–19. https://doi.org/10.1093/analys/58.1.7.
Clark, Andy, Karl Friston, and Sam Wilkinson. 2019. "Bayesing Qualia: Consciousness as Inference, Not Raw Datum." *Journal of Consciousness Studies* 26 (9–10): 19–33.
Clark, Robert E., and Larry R. Squire. 1998. "Classical Conditioning and Brain Systems: The Role of Awareness." *Science* 280 (5360): 77–81. https://doi.org/10.1126/science.280.5360.77.
Clark, Robert E., and Larry R. Squire. 1999. "Human Eyeblink Classical Conditioning: Effects of Manipulating Awareness of the Stimulus Contingencies." *Psychological Science* 10 (1): 14–18. https://doi.org/10.1111/1467-9280.00099.
Cleeremans, Axel, and Catherine Tallon-Baudry. 2022. "Consciousness Matters: Phenomenal Experience Has Functional Value." *Neuroscience of Consciousness* 2022 (1): niac007. https://doi.org/10.1093/nc/niac007.
Cochrane, Tom. 2021. "A Case of Shared Consciousness." *Synthese* 199 (1–2): 1019–37. https://doi.org/10.1007/s11229-020-02753-6.
Cohen, Michael A., and Daniel C. Dennett. 2011. "Consciousness Cannot Be Separated from Function." *Trends in Cognitive Sciences* 15 (8): 358–64. https://doi.org/10.1016/j.tics.2011.06.008.

Cohen, Stewart. 1984. "Justification and Truth." *Philosophical Studies* 46 (3): 279–95. https://doi.org/10.1007/BF00372907.

Colaço, David. 2022. "What Counts as a Memory? Definitions, Hypotheses, and 'Kinding in Progress.'" *Philosophy of Science* 89 (1): 89–106. https://doi.org/10.1017/psa.2021.14.

Coleman, Sam. 2019. "Natural Acquaintance." In *New Essays on Acquaintance*, edited by Jonathan Knowles and Thomas Raleigh. Oxford University Press. https://doi.org/10.1093/oso/9780198803461.003.0003.

Columbus, Christopher. 2010. "Journal of the First Voyage of Columbus." In *Journal of Christopher Columbus (During His First Voyage, 1492–93)*, 1st ed., edited by Clements R. Markham. Cambridge University Press. https://doi.org/10.1017/CBO9780511708411.003.

Comesaña, Juan. 2002. "The Diagonal and the Demon." *Philosophical Studies* 110 (3): 249–66. JSTOR. https://doi.org/10.1023/A:1020656411534.

Comte, Auguste. 1989. *Cours de Philosophie Positive: Texte Intégral*. Edited by Patrick Dupouey. Nathan.

Coninx, Sabrina. 2024. "Pain Philosophy: Recent Debates and Future Challenges." *Philosophy Compass* 19 (4): e12981. https://doi.org/10.1111/phc3.12981.

Constantinescu, Cristian. 2014. "Moral Vagueness: A Dilemma for Non-Naturalism." In *Oxford Studies in Metaethics: Volume 9*, edited by Russ Shafer-Landau. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198709299.003.0007.

Cornman, James W. 1968a. "Mental Terms, Theoretical Terms, and Materialism." *Philosophy of Science* 35 (1): 45–63. https://doi.org/10.1086/288186.

Cornman, James W. 1968b. "On the Elimination of 'Sensations' and Sensations." *The Review of Metaphysics* 22 (1): 15–35.

Cory, Therese Scarpelli. 2014. *Aquinas on Human Self-Knowledge*. Cambridge University Press. https://doi.org/10.1017/CBO9781107337619.

Cory, Therese Scarpelli. 2021. "Consciousness." In *The Routledge Companion to Medieval Philosophy*, edited by Cross, Richard and J.T. Paasch. Routledge, Taylor & Francis Group.

Cournot, Antoine-Augustin. 1851. *Essai Sur Les Fondements de Nos Connaissances et Sur Les Caractères de La Critique Philosophie*. Vol. 2. Hachette.

Cova, Florian. 2023. "A Defense of Natural Compatibilism." In *A Companion to Free Will*, 1st ed., edited by Joseph Campbell, Kristin M. Mickelson, and V. Alan White. Wiley. https://doi.org/10.1002/9781119210177.ch18.

Cova, Florian, Maxence Gaillard, and François Kammerer. 2021. "Is the Phenomenological Overflow Argument Really Supported by Subjective Reports?" *Mind and Language* 36 (3): 422–50. https://doi.org/10.1111/mila.12291.

Crick, Francis, and Christof Koch. 1990. "Towards a Neurobiological Theory of Consciousness." *Seminars in the Neurosciences* 2: 263–75.

Crisp, Roger. 2006. *Reasons and the Good*. Oxford University Press.

Crisp, Roger. 2021. "Well-Being." In *The Stanford Encyclopedia of Philosophy*, Winter 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/well-being/.

Cutter, Brian. 2017. "The Metaphysical Implications of the Moral Significance of Consciousness." *Philosophical Perspectives* 31 (1): 103–30. https://doi.org/10.1111/phpe.12092.

Cutter, Brian. 2022. "The Inconceivability Argument." *Ergo: An Open Access Journal of Philosophy* 9 (12): 329–56. https://doi.org/10.3998/ergo.2268.

Cutter, Brian, and Dustin Crummett. 2025. "Psychophysical Harmony: A New Argument for Theism." In *Oxford Studies in Philosophy of Religion*, edited by Lara Buchak and Dean W Zimmerman, vol. 11. Oxford University Press. https://doi.org/10.1093/9780198954712.003.0002.

Dagron, Tristan. 2009. *Toland et Leibniz. L'invention du néo-spinozisme*. Bibliothèque d'Histoire de la Philosophie. Vrin. Cairn.info. https://doi.org/10.3917/vrin.dagro.2009.01.

Dawkins, Marian Stamp. 2021. *The Science of Animal Welfare: Understanding What Animals Want*. Oxford University Press. https://doi.org/10.1093/oso/9780198848981.001.0001.

Dehaene, Stanislas. 2014. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Penguin books.
Dehaene, Stanislas, Hakwan Lau, and Sid Kouider. 2017. "What Is Consciousness, and Could Machines Have It?" *Science* 358 (6362): 486–92. https://doi.org/10.1126/science.aan8871.
Dehaene, Stanislas, and Lionel Naccache. 2001. "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework." *Cognition* 79: 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2.
Deijl, Willem van der. 2019. "Is Pleasure All That Is Good about Experience?" *Philosophical Studies* 176: 1769–87. https://doi.org/10.1007/s11098-018-1090-y.
Deijl, Willem van der. 2021. "The Sentience Argument for Experientialism about Welfare." *Philosophical Studies* 178 (1): 187–208. https://doi.org/10.1007/s11098-020-01427-w.
Della Rocca, Michael. 2008. *Spinoza*. Routledge Philosophers. Routledge.
Delon, Nicolas. 2024. "Strangers to Ourselves: A Nietzschean Challenge to the Badness of Suffering." *Inquiry: An Interdisciplinary Journal of Philosophy* 67 (9): 3600–3629. https://doi.org/10.1080/0020174X.2021.2018357.
Delon, Nicolas. manuscript. *Against Sentientism*. https://nicolasdelon.com/resources/Agential-value-Rome.pdf.
Demircioğlu, Erhan. 2013. "Physicalism and Phenomenal Concepts." *Philosophical Studies* 165 (1): 257–77. https://doi.org/10.1007/s11098-012-9959-7.
Dennett, Daniel. 1979. "On the Absence of Phenomenology." In *Body, Mind, and Method: Essays in Honor of Virgil C. Aldrich*, edited by Donald F. Gustafson and Bangs L. Tapscott. Synthese Library. Springer Netherlands. https://doi.org/10.1007/978-94-009-9479-9_6.
Dennett, Daniel. 1987. "Consciousness." In *Oxford Companion to the Mind*, edited by R.L. Gregory and O.L. Zangwill. Oxford University Press.
Dennett, Daniel. 1988. "Quining Qualia." In *Consciousness in Contemporary Science*, edited by Anthony Marcel and Eduardo Bisiach. Oxford University Press.
Dennett, Daniel. 1991. *Consciousness Explained*. Penguin.
Dennett, Daniel. 1995a. "Animal Consciousness: What Matters and Why." *Social Research* 62 (3): 691–710.
Dennett, Daniel. 1995b. "Cognitive Science as Reverse Engineering Several Meanings of 'Top-down' and 'Bottom-Up.'" In *Studies in Logic and the Foundations of Mathematics*, vol. 134, edited by Dag Prawitz, Brian Skyrms, and Dag Westerståhl. Elsevier.
Dennett, Daniel. 1998. "The Myth of Double Transduction." In *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, edited by Stuart R. Hameroff, Alfred W. Kaszniak, and Alwyn Scott, vol. 2. MIT press.
Dennett, Daniel. 2006. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness* . 1. paperback ed., 4. print. The Jean Nicod Lectures. MIT Press.
Dennett, Daniel. 2016. "Illusionism as the Obvious Default Theory of Consciousness." *Journal of Consciousness Studies* 23 (11–12): 65–72.
Dennett, Daniel. 2018. "The Fantasy of First-Person Science." In *The Map and the Territory: Exploring the Foundations of Science, Thought and Reality*, edited by Wuppuluri Shyam and Francisco Antonio Dorio. Springer.
Dennett, Daniel. 2019. "Welcome to Strong Illusionism." *Journal of Consciousness Studies* 26 (9–10): 48–58.
Denton, D.A., M.J. McKinley, M. Farrell, and G.F. Egan. 2009. "The Role of Primordial Emotions in the Evolutionary Origin of Consciousness." *Consciousness and Cognition* 18 (2): 500–514. https://doi.org/10.1016/j.concog.2008.06.009.
Descartes, René. 1985a. *The Philosophical Writings of Descartes*. 1st ed. Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch. Vol. 1. Cambridge University Press. https://doi.org/10.1017/CBO9780511805042.
Descartes, René. 1985b. *The Philosophical Writings of Descartes*. 1st ed. Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch. Vol. 2. Cambridge University Press. https://doi.org/10.1017/CBO9780511805042.

Díaz, Rodrigo. 2021. "Do People Think Consciousness Poses a Hard Problem?: Empirical Evidence on the Meta-Problem of Consciousness." *Journal of Consciousness Studies* 28 (3–4): 55–75.

Dorsey, Dale. 2010. "Three Arguments for Perfectionism." *Noûs* 44 (1): 59–79. https://doi.org/10.1111/j.1468-0068.2009.00731.x.

Dorsey, Dale. 2021. *A Theory of Prudence*. 1st[st] ed. Oxford University Press.

Dostoevsky, Fyodor. [1880] 1992. *The Brothers Karamazov*. 1st ed. Translated by Richard Pevear and Larissa Volokhonsky. Vintage Classics.

Dretske, Fred. 1981. *Knowledge and the Flow of Information*. MIT Press.

Dretske, Fred. 1988. *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

Dretske, Fred. 1995. *Naturalizing the Mind*. MIT Press.

Dretske, Fred. 2003. "Experience as Representation." *Philosophical Issues* 13 (1): 67–82.

Duncan, Matt. 2021. "Acquaintance." *Philosophy Compass* 16 (3). https://doi.org/10.1111/phc3.12727.

Duncan, Matt. 2023. "How You Know You're Conscious: Illusionism and Knowledge of Things." *Review of Philosophy and Psychology* 14 (1): 185–205. https://doi.org/10.1007/s13164-021-00590-1.

Dung, Leonard. 2022. "Does Illusionism Imply Skepticism of Animal Consciousness?" *Synthese* 200 (3): 238. https://doi.org/10.1007/s11229-022-03710-1.

Dung, Leonard. 2024. "Preserving the Normative Significance of Sentience." *Journal of Consciousness Studies* 31 (1–2): 8–30. https://doi.org/10.53765/20512201.31.1.008.

Dung, Leonard. 2025. "Track Record Arguments in Normative Ethics." *Pacific Philosophical Quarterly* n/a (n/a). https://doi.org/10.1111/papq.70003.

Dung, Leonard, and Albert Newen. 2023. "Profiles of Animal Consciousness: A Species-Sensitive, Two-Tier Account to Quality and Distribution." *Cognition* 235 (June): 105409. https://doi.org/10.1016/j.cognition.2023.105409.

Dung, Leonard and François Kammerer. 2025. "A Science of Chimeras? The Implications of Illusionism for Non-human Consciousness Research." *Philosophical Psychology*: 1–2410. https://doi.org/1080/09515089.2025.2587227.

Dworkin, Ronald. 2011. *Justice for Hedgehogs*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vkt.

Ericsson, Anders. 2003. "Valid and Non-Reactive Verbalization of Thoughts during Performance of Tasks towards a Solution to the Central Problems of Introspection as a Source of Scientific Data." *Journal of Consciousness Studies* 10 (9–10): 1–18.

Esfeld, Michael. 2009. "The Modal Nature of Structures in Ontic Structural Realism." *International Studies in the Philosophy of Science* 23 (2): 179–94. https://doi.org/10.1080/02698590903006917.

European Union. 2007. "EUR-Lex - 12007L/TXT - EN - EUR-Lex." https://eur-lex.europa.eu/eli/treaty/lis/sign.

Farrell, B.A. 1950. "Experience." *Mind* 59 (April): 170–98.

Feest, Uljana. 2014. "Phenomenal Experiences, First-Person Methods, and the Artificiality of Experimental Data." *Philosophy of Science* 81 (5): 927–39. https://doi.org/10.1086/677689.

Feldman, Fred. 2002. "The Good Life: A Defense of Attitudinal Hedonism." *Philosophy and Phenomenological Research* 65: 604–27. https://doi.org/10.1111/j.1933-1592.2002.tb00223.x.

Feyerabend, Paul. 1963. "Mental Events and the Brain." *Journal of Philosophy* 60 (11): 295–96.

Feyerabend, Paul. 1993. *Against Method*. 3rd ed. Verso.

Fine, Kit. 1995. "Senses of Essence." In *Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus*, edited by Walter Sinnott-Armstrong, Diana Raffman, and Nicholas Asher. Cambridge University Press.

Fine, Kit. 2002. "The Question of Realism." In *Individuals, Essence and Identity*, edited by Ermanno Bencivenga, vol. 4, edited by Andrea Bottani, Massimiliano Carrara, and

Pierdaniele Giaretta. Topoi Library. Springer Netherlands. https://doi.org/10.1007/978-94-017-1866-0_1.
Fish, William. 2008. "Relationalism and the Problems of Consciousness." *Teorema: Revista Internacional de Filosofía* 27 (3): 167–80.
Fisher, Saul. 2024. "Pierre Gassendi." In *The Stanford Encyclopedia of Philosophy*, Spring 2024, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2024/entries/gassendi/.
Fisk, John. 2004. "Conjunction Fallacy." In *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgment and Memory*, edited by Rüdiger Pohl. Psychology Press.
Flavell, John H. 1979. "Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry." *American Psychologist* 34 (10): 906–11. https://doi.org/10.1037/0003-066X.34.10.906.
Fleming, Stephen M. 2021. *Know Thyself: The Science of Self-Awareness*. 1st ed. Basic Books.
Fleming, Stephen M, Chris D. Frith, Mel Goodale, et al. 2023. *The Integrated Information Theory of Consciousness as Pseudoscience*. Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/zsr78.
Fletcher, Guy. 2013. "A Fresh Start for the Objective-List Theory of Well-Being." *Utilitas* 25 (2): 206–20. https://doi.org/10.1017/S0953820812000453.
Fletcher, Guy. 2017. "Pain for the Moral Error Theory? A New Companions-in-Guilt Argument." *Australasian Journal of Philosophy* 96 (3): 474–82. https://doi.org/10.1080/00048402.2017.1386693.
Fletcher, Guy. 2022. "A Painful End for Perfectionism?" *Midwest Studies in Philosophy* 46: 233–50. https://doi.org/10.5840/msp2022112328.
Fodor, Jerry. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
Fodor, Jerry. 2004. Review of *You Can't Argue with a Novel*, by Dan Lloyd. *London Review of Books*, March 4. https://www.lrb.co.uk/the-paper/v26/n05/jerry-fodor/you-can-t-argue-with-a-novel.
Fodor, Jerry A. 1990. *A Theory of Content and Other Essays*. MIT Press.
Forrester, Paul. 2024. "The Desire Machine." *Analysis* 84 (2): 249–57. https://doi.org/10.1093/analys/anad061.
Frances, Bryan. 2008. "Live Skeptical Hypotheses." In *The Oxford Handbook of Skepticism*, edited by John Greco. Oxford University Press.
Francken, Jolien C, Lola Beerendonk, Dylan Molenaar, et al. 2022. "An Academic Survey on Theoretical Foundations, Common Assumptions and the Current State of Consciousness Science." *Neuroscience of Consciousness* 2022 (1): 1–13. https://doi.org/10.1093/nc/niac011.
Franco, Eli. 1997. *Dharmakīrti on Compassion and Rebirth*. Arbeitkreis für Tibetische und Buddhistische Studien Universität Wien.
Frankfurt, Harry. 1982. "The Importance of What We Care About." *Synthese* 53 (2): 257–72. https://doi.org/10.1007/BF00484902.
Frankfurt, Harry. 1998. "Some Thoughts about Caring." *Ethical Perspectives*, no. 1: 3–14. https://doi.org/10.2143/EP.5.1.563101.
Frankish, Keith. 2007. "The Anti-Zombie Argument." *Philosophical Quarterly* 57 (229): 650–66. https://doi.org/10.1111/j.1467-9213.2007.510.x.
Frankish, Keith. 2012. "Quining Diet Qualia." *Consciousness and Cognition* 21 (2): 667–76. https://doi.org/10.1016/j.concog.2011.04.001.
Frankish, Keith. 2016. "Illusionism as a Theory of Consciousness." *Journal of Consciousness Studies* 23 (11–12): 11–39.
Frankish, Keith. 2019. "The Meta-Problem Is The Problem of Consciousness." *Journal of Consciousness Studies* 26 (9–10): 9–10.
Frankish, Keith. 2023. "What Is Illusionism?" *Klesis* 55: 1–13.
Frankish, Keith. 2024. "The Ethical Implications of Illusionism." *Neuroethics* 17 (2): 28. https://doi.org/10.1007/s12152-024-09562-5.

French, Steven, and James Ladyman. 2010. "In Defence of Ontic Structural Realism." In *Scientific Structuralism*, edited by Alisa Bokulich and Peter Bokulich, vol. 281. Boston Studies in the Philosophy and History of Science. Springer Netherlands. https://doi.org/10.1007/978-90-481-9597-8_2.

Fumerton, Richard. 1995. *Metaepistemology and Skepticism*. Rowman and Littlefield.

Fumerton, Richard. 2016. "The Prospects for Traditional Internalism." In *Intellectual Assurance*, edited by Brett Coppenger and Michael Bergmann. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198719632.003.0013.

Galvin, Susan J., John V. Podd, Vit Drga, and John Whitmore. 2003. "Type 2 Tasks in the Theory of Signal Detectability: Discrimination between Correct and Incorrect Decisions." *Psychonomic Bulletin & Review* 10 (4): 843–76. https://doi.org/10.3758/BF03196546.

Ganson, Todd Stuart. 1997. "What's Wrong with the Aristotelian Theory of Sensible Qualities?" *Phronesis* 42 (3): 263–82.

Garfield, Jay L. 2016. "Illusionism and Givenness." *Journal of Consciousness Studies* 23 (11–12): 73–82.

Gauvry, Charlotte, and Rüber Theodor. 2024. "Extrapolating Consciousness in Isolated Hemispheres. Hemispherotomy as a New Challenge." *Revue de Métaphysique et de Morale* 1 (121): 47–65. https://doi.org/10.3917/rmm.241.0047.

Gertler, Brie. 2012. "Renewed Acquaintance." In *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744794.003.0004.

Giustina, Anna. 2022. "An Acquaintance Alternative to Self-Representationalism." *Philosophical Studies* 179 (12): 3831–63. https://doi.org/10.1007/s11098-022-01868-5.

Glannon, Walter. 2016. "The Value and Disvalue of Consciousness." *Cambridge Quarterly of Healthcare Ethics* 25 (4): 600–612. https://doi.org/10.1017/S0963180116000335.

Glidden, David K. 1979. "Epicurus on Self-Perception." *American Philosophical Quarterly* 16 (4): 297–306. JSTOR.

Goff, Philip. 2015. "Real Acquaintance and Physicalism." In *Phenomenal Qualities: Sense, Perception and Consciousness*, edited by Paul Coates and Sam Coleman. Oxford University Press. https://academic.oup.com/book/11696/chapter/160654061.

Goff, Philip. 2016. "Is Realism about Consciousness Compatible with a Scientifically Respectable Worldview?" *Journal of Consciousness Studies* 23 (11–12): 83–97.

Goff, Philip. 2017. *Consciousness and Fundamental Reality*. Oxford University Press.

Goldman, Alvin. 1979. "What Is Justified Belief?" In *Justification and Knowledge*, edited by George Pappas. D. Reidel.

Goldman, Alvin. 1993. "Consciousness, Folk Psychology, and Cognitive Science." *Consciousness and Cognition* 2 (4): 364–82.

Goldman, Alvin, and Bob Beddor. 2021. "Reliabilist Epistemology." In *The Stanford Encyclopedia of Philosophy*, Summer 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/reliabilism/.

Graziano, Michael. 2013. *Consciousness and the Social Brain*. Oxford University Press.

Graziano, Michael. 2016. "Consciousness Is Not Mysterious." *The Atlantic*, January 12. http://www.theatlantic.com/science/archive/2016/01/consciousness-color-brain/423522/.

Graziano, Michael. 2019a. *Rethinking Consciousness: A Scientific Theory of Subjective Experience*. W.W. Norton & Company.

Graziano, Michael. 2019b. "We Are Machines That Claim to Be Conscious." *Journal of Consciousness Studies* 26 (9–10): 9–10.

Graziano, Michael. 2024. "Illusionism Big and Small: Some Options for Explaining Consciousness." *Eneuro* 11 (10). https://doi.org/10.1523/ENEURO.0210-24.2024.

Graziano, Michael, Arvid Guterstam, Branden J. Bio, and Andrew I. Wilterson. 2020. "Toward a Standard Model of Consciousness: Reconciling the Attention Schema, Global Workspace, Higher-Order Thought, and Illusionist Theories." *Cognitive Neuropsychology* 37 (3–4): 155–72. https://doi.org/10.1080/02643294.2019.1670630.

Gregory, Daniel, Malte Hendrickx, and Cameron Turner. 2022. "Who Knows What Mary Knew? An Experimental Study." *Philosophical Psychology* 35 (4): 522–45. https://doi.org/10.1080/09515089.2021.2001448.

Griesinger, Wilhelm. 1861. *Pathologie Und Therapie Der Psychischen Krankheiten*. 2nd ed. Braunschweig.

Gruen, Lori. 2011. *Ethics and Animals: An Introduction*. Cambridge Applied Ethics. Cambridge University Press. https://doi.org/10.1017/CBO9780511976162.

Gupta, Anil. 2006. *Empiricism and Experience*. Oxford University Press.

Güzeldere, Güven. 1997. "The Many Faces of Consciousness: A Field Guide." In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere. MIT Press.

Hall, Richard. 2007. "Phenomenal Properties as Dummy Properties." *Philosophical Studies* 135 (2): 2. https://doi.org/10.1007/s11098-005-1064-8.

Hameroff, Stuart, and Roger Penrose. 2014. "Consciousness in the Universe: A Review of the 'Orch OR' Theory." *Physics of Life Reviews* 11 (1): 39–78. https://doi.org/10.1016/j.plrev.2013.08.002.

Hamlin, J. Kiley, Karen Wynn, and Paul Bloom. 2010. "Three-month-olds Show a Negativity Bias in Their Social Evaluations." *Developmental Science* 13 (6): 923–29. https://doi.org/10.1111/j.1467-7687.2010.00951.x.

Hanna, Robert. 2016. "Conceptual Analysis." In *Routledge Encyclopedia of Philosophy*, 1st ed. Routledge. https://doi.org/10.4324/9780415249126-U033-1.

Harman, Gilbert. 1973. *Thought*. Princeton University Press.

Harman, Gilbert. 1987. "(Non-Solipsistic) Conceptual Role Semantics." In *New Directions in Semantics*, edited by Ernest Lepore. Academic Press.

Harman, Gilbert. 2003. "Skepticism and Foundations." In *The Skeptics: Contemporary Essays*, edited by Steven Luper. Ashgate.

Hasan, Ali, and Richard Fumerton. 2022. "Foundationalist Theories of Epistemic Justification." In *The Stanford Encyclopedia of Philosophy*, Fall 2022, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2022/entries/justep-foundational/.

Hatfield, Gary. 2005. "Introspective Evidence in Psychology." In *Scientific Evidence: Philosophical Theories & Applications*, edited by P. Achinstein. The Johns Hopkins University Press.

Heathwood, Chris. 2006. "Desire Satisfactionism and Hedonism." *Philosophical Studies* 128: 539–63. https://doi.org/10.1007/s11098-004-7817-y.

Heathwood, Chris. 2011. "Desire-Based Theories of Reasons, Pleasure, and Welfare." In *Oxford Studies in Metaethics: Volume 6*, edited by Russ Shafer-Landau. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199606375.003.0004.

Heathwood, Chris. 2019. "Which Desires Are Relevant to Well-Being?" *Noûs* 53 (3): 664–88. https://doi.org/10.1111/nous.12232.

Hempel, Carl G. 1980. "Comments on Goodman's Ways of Worldmaking." *Synthese* 45 (2): 193–99.

Hetherington, Stephen. 2024. "Fallibilism." In *Internet Encyclopedia of Philosophy*. https://iep.utm.edu.

Hill, Christopher. 1991. *Sensations: A Defense of Type Materialism*. Cambridge University Press.

Hill, Christopher. 1997. "Imaginability, Conceivability, Possibility and the Mind-Body Problem." *Philosophical Studies* 87: 61–85.

Hill, Christopher. 2009. *Consciousness*. Cambridge University Press.

Horgan, John. 2014. "Physics Titan Still Thinks String Theory Is 'On the Right Track.'" Scientific American Blog Network. https://blogs.scientificamerican.com/cross-check/physics-titan-still-thinks-string-theory-is-on-the-right-track/.

Horgan, John. 2016. "World's Smartest Physicist Thinks Science Can't Crack Consciousness." Scientific American Blog Network. https://blogs.scientificamerican.com/cross-check/world-s-smartest-physicist-thinks-science-can-t-crack-consciousness/.

Horgan, Terry. 2012. "Introspection about Phenomenal Consciousness: Running the Gamut from Infallibility to Impotence." In *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744794.003.0015.

Huemer, Michael. 2001. *Skepticism and the Veil of Perception*. Lanham.

Huemer, Michael. 2007. "Compassionate Phenomenal Conservatism." *Philosophy and Phenomenological Research* 74 (1): 1. https://doi.org/10.1111/j.1933-1592.2007.00002.x.

Huemer, Michael. 2013. "Phenomenal Conservatism Über Alles." In *Seemings and Justification*, edited by Chris Tucker. Oxford University Press. https://academic.oup.com/book/4517/chapter/146597813.

Hume, David. [1739–1740] 1896. *A Treatise of Human Nature*. Edited by Lewis Amherst Selby-Bigge. Oxford, Clarendon press.

Humphrey, Nicholas. 2011. *Soul Dust: The Magic of Consciousness*. Princeton University Press.

Humphrey, Nicholas. 2016. "Redder than Red: Illusionism or Phenomenal Surrealism?" *Journal of Consciousness Studies* 23 (11–12): 35–55.

Humphrey, Nicholas. 2019. "Easy Does It: A Soft Landing for Consciousness." *Journal of Consciousness Studies* 26 (9–10): 105–14.

Hurlburt, Russell T., and Christopher L. Heavey. 2001. "Telling What We Know: Describing Inner Experience." *Trends in Cognitive Sciences* 5 (9): 400–403. https://doi.org/10.1016/S1364-6613(00)01724-1.

Husserl, Edmund. [1931] 1960. *Cartesian Meditations*. Translated by Dorion Cairns. Martinus Nijhoff.

Husserl, Edmund. [1910–1911] 2002. "Philosophy as Rigorous Science." Translated by Marcus Brainard. *The New Yearbook for Phenomenology and Phenomenological Philosophy* II: 249–95.

Husserl, Edmund. [1900–1901] 2013. *Logical Investigations Volume 2*. Translated by J.N. Findlay. Routledge.

Huxley, Thomas. 1866. *Lessons in Elementary Physiology*. Macmillan.

Ibn Tufayl, Muhammad ibn 'Abd al-Malik. 2009. *Ibn Tufayl's Hayy Ibn Yaqzān: A Philosophical Tale*. [5th ed.] updated 2009. Translated by Lenn Evan Goodman. University of Chicago Press.

Impagnatiello, Michele Odisseas. forthcoming. "Ethics Vs. Metaphysics." *Journal of Philosophy*.

Irvine, Elizabeth. 2012. *Consciousness as a Scientific Concept: A Philosophy of Science Perspective*. Vol. 5. Springer Science & Business Media.

Irvine, Elizabeth. 2013. "Measures of Consciousness." *Philosophy Compass* 8 (3): 285–97. https://doi.org/10.1111/phc3.12016.

Irvine, Elizabeth. 2021. "Developing Dark Pessimism towards the Justificatory Role of Introspective Reports." *Erkenntnis* 86: 1319–44. https://doi.org/10.1007/s10670-019-00156-9.

Irvine, Elizabeth, and Mark Sprevak. 2020. "Eliminativism about Consciousness." In *Oxford Handbook of the Philosophy of Consciousness*, edited by Uriah Kriegel. Oxford University Press.

Jackson, Frank. 1977. *Perception*. Cambridge University Press.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32 (April): 127–36.

Jackson, Frank. 1986. "What Mary Didn't Know." *Journal of Philosophy* 83 (May): 291–95.

Jackson, Frank. 1998a. *From Metaphysics to Ethics*. Clarendon Press.

Jackson, Frank. 1998b. "Postscript on Qualia." In *Mind, Methods and Conditionals*, 76–79. Routledge.

Jacob, François. 1977. "Evolution and Tinkering." *Science* 196 (4295): 1161–66.

Jaquet, François. 2019. "Sorting Out Solutions to the Now-What Problem." *Journal of Ethics and Social Philosophy* 17 (3): 239–58. https://doi.org/10.26556/jesp.v17i3.812.

Jensen, Karsten Klint, and Peter Sandøe. 1997. "Animal Welfare: Relative or Absolute?" *Applied Animal Behaviour Science* 54 (1): 33–37. https://doi.org/10.1016/S0168-1591(96)01203-8.

Joyce, Richard. 2005. "Moral Fictionalism." In *Fictionalism in Metaphysics*, edited by Mark Eli Kalderon. Oxford University Press UK.

Kagan, Shelly. 2019. *How to Count Animals, More or Less*. 1st ed. Uehiro Series in Practical Ethics. Oxford University Press.
Kahane, Guy. 2009. "Pain, Dislike and Experience." *Utilitas* 21 (3): 327–36. https://doi.org/10.1017/S0953820809990070.
Kammerer, François. 2015. "How a Materialist Can Deny That the United States Is Probably Conscious—Response to Schwitzgebel." *Philosophia* 43 (4): 1047–57. https://doi.org/10.1007/s11406-015-9653-z.
Kammerer, François. 2016a. "Conscious Experiences as Ultimate Seemings: Renewing the Phenomenal Concept Strategy." *Argumenta* 1 (2): 233–43.
Kammerer, François. 2016b. "Le problème de l'expérience consciente: une tentative de dissolution." PhD dissertation, Université Paris-Sorbonne.
Kammerer, François. 2016c. "The Hardest Aspect of the Illusion Problem—and How to Solve It." *Journal of Consciousness Studies* 23 (11–12): 123–39.
Kammerer, François. 2018a. "Can You Believe It? Illusionism and the Illusion Meta-Problem." *Philosophical Psychology* 31 (1): 44–67. https://doi.org/10.1080/09515089.2017.1388361.
Kammerer, François. 2018b. "Is the Antipathetic Fallacy Responsible for the Intuition That Consciousness Is Distinct from the Physical?" *Croatian Journal of Philosophy* 18 (1): 59–73.
Kammerer, François. 2019a. *Conscience et Matière. Une Solution Matérialiste Au Problème de l'expérience Consciente*. Editions Matériologiques.
Kammerer, François. 2019b. "Does the Explanatory Gap Rest on a Fallacy?" *Review of Philosophy and Psychology* 10: 649–67. https://doi.org/10.1007/s13164-018-0424-1.
Kammerer, François. 2019c. "The Meta-Problem of Consciousness and the Evidential Approach." *Journal of Consciousness Studies* 26 (9–10): 124–35.
Kammerer, François. 2020. "The Normative Challenge for Illusionist Views of Consciousness." *Ergo* 6 (32): 891–924. https://doi.org/10.3998/ergo.12405314.0006.032.
Kammerer, François. 2021. "The Illusion of Conscious Experience." *Synthese* 198: 845–66. https://doi.org/10.1007/s11229-018-02071-y.
Kammerer, François. 2022a. "Ethics without Sentience: Facing up to the Probable Insignificance of Phenomenal Consciousness." *Journal of Consciousness Studies* 29 (3–4): 180–204. https://doi.org/10.53765/20512201.29.3.180.
Kammerer, François. 2022b. "How Rich Is the Illusion of Consciousness?" *Erkenntnis* 87 (2): 499–515. https://doi.org/10.1007/s10670-019-00204-4.
Kammerer, François. 2022c. "How Can You Be So Sure? Illusionism and the Obviousness of Phenomenal Consciousness." *Philosophical Studies* 179 (9): 2845–2867. https://doi.org/10.1007/s11098-022-01804-7.
Kammerer, François. 2023. "Certainty and Our Sense of Acquaintance with Experiences." *Erkenntnis* 88: 3015–36. https://doi.org/10.1007/s10670-021-00488-5.
Kammerer, François. 2024. "Sentientism Still under Threat—Reply to Dung." *Journal of Consciousness Studies* 31 (3): 103–19. https://doi.org/10.53765/20512201.31.3.103.
Kammerer, François. 2025. "Defining Consciousness and Denying Its Existence: Sailing between Charybdis and Scylla." *Philosophical Studies* 182 (2): 541–65. https://doi.org/10.1007/s11098-025-02285-0.
Kammerer, François, and Keith Frankish. 2023a. "More Possibilities for Introspection Reply to Commentators." *Journal of Consciousness Studies* 30 (9): 235–75. https://doi.org/10.53765/20512201.30.9.235.
Kammerer, François, and Keith Frankish. 2023b. "What Forms Could Introspective Systems Take? A Research Programme." *Journal of Consciousness Studies* 30 (9): 13–48. https://doi.org/10.53765/20512201.30.9.013.
Kastrup, Bernardo. 2018. "The Universe in Consciousness." *Journal of Consciousness Studies* 25 (5–6): 125–55.
Kellner, Birgit. 2014. "Changing Frames in Buddhist Thought: The Concept of Ākāra in Abhidharma and in Buddhist Epistemological Analysis." *Journal of Indian Philosophy* 42 (2–3): 275–95. https://doi.org/10.1007/s10781-013-9190-7.

Kellner, Birgit. 2015. "The Oxford Handbook of Indian Philosophy." In *Proving Idealism. Dharmakīrti*, edited by Jonardon Ganeri. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199314621.013.18.

Kelly, Thomas. 2005. "Moorean Facts and Belief Revision, or Can the Skeptic Win?" *Philosophical Perspectives* 19 (1): 179–209. https://doi.org/10.1111/j.1520-8583.2005.00059.x.

Kelly, Thomas. 2008. "Common Sense as Evidence: Against Revisionary Ontology and Skepticism." *Midwest Studies in Philosophy* 32 (1): 53–78. https://doi.org/10.1111/j.1475-4975.2008.00165.x.

Kendall, David. 1968. "Obituary: A. C. Aitken, D.Sc, F.R.S." *Proceedings of the Edinburgh Mathematical Society* 16 (2): 151–76. https://doi.org/10.1017/S0013091500012530.

Kind, Amy. 2001. "Qualia Realism." *Philosophical Studies* 104 (2): 143–62. https://doi.org/10.1023/A:1010317611694.

King, Peter. 2007a. "Rethinking Representation in the Middle Ages." In *Representations and Objects of Thought in Medieval Philosophy*, 1st ed., edited by Henrik Lagerlund. Routledge.

King, Peter. 2007b. "Why Isn't the Mind-Body Problem Medieval?" In *Forming the Mind: Essays on the Internal Senses and the Mind/Body Problem from Avicenna to the Medical Enlightenment*, edited by Henrik Lagerlund. Springer.

Kirk, Robert. 1974. "Sentience and Behaviour." *Mind* 83: 43–60.

Klein, Colin, and Andrew B. Barron. 2016. "Insects Have the Capacity for Subjective Experience." *Animal Sentience* 1 (9): 1–19. https://doi.org/10.51291/2377-7478.1113.

Knobe, Joshua, and Shaun Nichols. 2017. "Experimental Philosophy." In *The Stanford Encyclopedia of Philosophy*, Winter 2017, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/.

Koch, Christof, Marcello Massimini, Melanie Boly, and Giulio Tononi. 2016. "Neural Correlates of Consciousness: Progress and Problems." *Nature Reviews Neuroscience* 17 (5): 307–21. https://doi.org/10.1038/nrn.2016.22.

Korman, Daniel Z. 2019. "Debunking Arguments." *Philosophy Compass* 14 (12). https://doi.org/10.1111/phc3.12638.

Korsgaard, Christine M. 2018. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford University Press. https://doi.org/10.1093/oso/9780198753858.001.0001.

Kouider, Sid, Jérôme Sackur, and Vincent De Gardelle. 2012. "Do We Still Need Phenomenal Consciousness? Comment on Block." *Trends in Cognitive Sciences* 16 (3): 140–41. https://doi.org/10.1016/j.tics.2012.01.003.

Kraut, Richard. 2022. "Aristotle's Ethics." In *The Stanford Encyclopedia of Philosophy*, Fall 2022, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/.

Kriegel, Uriah. 2005. "Naturalizing Subjective Character." *Philosophy and Phenomenological Research* 71: 23–56. https://doi.org/10.1111/j.1933-1592.2005.tb00429.x.

Kriegel, Uriah. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford University Press.

Kriegel, Uriah. 2015. *The Varieties of Consciousness*. Philosophy of Mind. Oxford University Press.

Kriegel, Uriah. 2018. *Brentano's Philosophical System*. Oxford University Press.

Kriegel, Uriah. 2019. "The Value of Consciousness." *Analysis* 79 (3): 503–20. https://doi.org/10.1093/analys/anz045.

Kriegel, Uriah. 2023. "The Structure of Phenomenal Justification." *Australasian Journal of Philosophy* 101 (2): 282–97. https://doi.org/10.1080/00048402.2021.1978513.

Kripke, Saul. 1980. *Naming and Necessity*. Harvard University Press.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. University of Chicago Press.

La Mettrie, Julien Offray de. 1996. *Machine Man and Other Writings*. Edited by Ann Thomson. Cambridge Texts in the History of Philosophy. Cambridge University Press.

Ladyman, James. 1998. "What Is Structural Realism?" *Studies in History and Philosophy of Science Part A* 29 (3): 409–24. https://doi.org/10.1016/S0039-3681(98)80129-5.

Lai, Emily R. 2011. "Metacognition: A Literature Review." *Always Learning: Pearson Research Report* 24: 1–40.

Lamme, Victor. 2006. "Towards a True Neural Stance on Consciousness." *Trends in Cognitive Sciences* 10 (11): 494–501. https://doi.org/10.1016/j.tics.2006.09.001.

Langsam, Harold L. 2011. *The Wonder of Consciousness: Understanding the Mind through Philosophical Reflection*. MIT Press.

Larkey, Levi B., and Arthur B. Markman. 2005. "Processes of Similarity Judgment." *Cognitive Science* 29 (6): 1061–76. https://doi.org/10.1207/s15516709cog0000_30.

Latham, Noa. 2000. "Chalmers on the Addition of Consciousness to the Physical World." *Philosophical Studies* 98 (1): 71–97.

Lau, Hakwan. 2022. *In Consciousness We Trust: The Cognitive Neuroscience of Subjective Experience*. Oxford University Press.

Lau, Hakwan, and David Rosenthal. 2011. "Empirical Support for Higher-Order Theories of Conscious Awareness." *Trends in Cognitive Sciences* 15 (8): 365–73. https://doi.org/10.1016/j.tics.2011.05.009.

Lau, Hawkan, and Matthias Michel. 2019. "A Socio-Historical Take on the Meta-Problem of Consciousness." *Journal of Consciousness Studies* 26 (9–10): 9–10.

Laurence, Stephen, and Eric Margolis. 1999. "Concepts and Cognitive Science." In *Concepts: Core Readings*, edited by Stephen Laurence and Eric Margolis. MIT Press.

Lavender, Jordan. 2022. "The Beatific Vision and the Metaphysics of Conscious Experience in John of Ripa." *Res Philosophica* 99 (2): 187–212. https://doi.org/10.11612/resphil.2155.

Lavender, Jordan. 2024. "Color, Qualia, and Subjectivity in Fourteenth-Century Scholasticism." Paper presented at Medieval Theories of Consciousness, Chicago. August 19.

Lebrecht, Sophie, Moshe Bar, Lisa Feldman Barrett, and Michael J. Tarr. 2012. "Micro-Valences: Perceiving Affective Valence in Everyday Objects." *Frontiers in Psychology* 3 (107). https://doi.org/10.3389/fpsyg.2012.00107.

Leduc, Christian, Paul Rateau, and Jean-Luc Solère, eds. 2015. *Leibniz et Bayle: confrontation et dialogue*. Philosophie 43. F. Steiner.

Lee, Andrew. 2018. "Is Consciousness Intrinsically Valuable?" *Philosophical Studies* 175 (1): 1–17. https://doi.org/10.1007/s11098-018-1032-8.

Lee, Andrew. 2025. "Consciousness Makes Things Matter." *Philosophers' Imprint* 25: 35. https://doi.org/10.3998/phimp.1956.

Lee, Andrew. forthcoming. "Metaethical Experientialism." In *The Importance of Being Conscious*, edited by Geoffrey Lee and Adam Pautz. Oxford University Press.

Lee, Geoffrey. 2014. "Materialism and the Epistemic Significance of Consciousness." In *Current Controversies in Philosophy of Mind*, edited by Uriah Kriegel. Routledge.

Lee, Sukjae. 2020. "Occasionalism." In *The Stanford Encyclopedia of Philosophy*, Fall 2020, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/occasionalism/.

Lehrer, Keith, and Stewart Cohen. 1983. "Justification, Truth, and Coherence." *Synthese* 55 (2): 191–207.

Leibniz, Gottfried Wilhelm. 1887. *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz*. Weidmann. Edited by Carl Immanuel Gerhardt. Vol. 3. Berlin.

Leibniz, Gottfried Wilhelm. 1989a. *Philosophical Essays*. Translated by Roger Ariew and Daniel Garber. Hackett.

Leibniz, Gottfried Wilhelm. [1714] 1989b. "The Principles of Nature and of Grace, Based on Reason." In *Philosophical Papers and Letters*, edited by Gottfried Wilhelm Leibniz and Leroy E. Loemker. Springer Netherlands. https://doi.org/10.1007/978-94-010-1426-7_67.

Leibniz, Gottfried Wilhelm. [1765] 1996. *Leibniz: New Essays on Human Understanding*. 2nd ed. Edited by Peter Remnant and Jonathan Bennett. Cambridge University Press. https://doi.org/10.1017/CBO9781139166874.

Leibniz, Gottfried Wilhelm. 2011. *Leibniz and the Two Sophies: The Philosophical Correspondence*. With Sophie Electress Of Hanover, Queen Sophie Charlotte Of Prussia, and Lloyd Strickland. Iter Press.

Levin, Janet. 2007. "What Is a Phenomenal Concept?" In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter and Sven Walter. Oxford University Press.

Levine, Joseph. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64 (October): 354–61.

Levine, Joseph. 1994. "Out of the Closet: A Qualophile Confronts Qualophobia." *Philosophical Topics* 22 (1/2): 107–26.

Levine, Joseph. 2001. *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.

Levine, Joseph. 2007. "Phenomenal Concepts and the Materialist Constraint." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter and Sven Walter. Oxford University Press.

Levy, Neil. 2014. "The Value of Consciousness." *Journal of Consciousness Studies* 21 (1–2): 127–38.

Levy, Neil. 2024. "Consciousness Ain't All That." *Neuroethics* 17 (2): 21. https://doi.org/10.1007/s12152-024-09559-0.

Levy, Neil, and Josh Shepherd. 2020. "Consciousness and Morality." In *The Oxford Handbook of the Philosophy of Consciousness*, edited by Uriah Kriegel. Oxford University Press.

Lewis, David. 1973a. "Causation." *Journal of Philosophy* 70 (17): 556. https://doi.org/10.2307/2025310.

Lewis, David. 1973b. *Counterfactuals*. Harvard University Press.

Lewis, David. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy* 62 (3): 221–36.

Lewis, David. 1995. "Should a Materialist Believe in Qualia?" *Australasian Journal of Philosophy* 73 (1): 140–44. https://doi.org/10.1080/00048409512346451.

Lin, Eden. 2020. "Attitudinal and Phenomenological Theories of Pleasure." *Philosophy and Phenomenological Research* 100 (3): 510–24. https://doi.org/10.1111/phpr.12558.

Lin, Eden. 2021. "The Experience Requirement on Well-Being." *Philosophical Studies* 178 (3): 867–86. https://doi.org/10.1007/s11098-020-01463-6.

Lindsay, D.S. 2008. "Source Monitoring." In *Learning and Memory: A Comprehensive Reference*, edited by John H. Byrne, vol. 2. Elsevier. https://doi.org/10.1016/B978-012370509-9.00175-3.

Liu, Haoying. 2020. "On Chalmers on the Meta-Problem." *Journal of Consciousness Studies* 27 (5–6): 91–98.

Liu, Michelle. 2021. "Revelation and the Intuition of Dualism." *Synthese* 199 (3–4): 3–4. https://doi.org/10.1007/s11229-021-03299-x.

Liu, Michelle. 2022. "A Critique of X-Phi on the Hard Problem of Consciousness." Ruhr-Universität Bochum. https://youtu.be/eQpfIWeKsv8.

Liu, Michelle. 2024. "Revelation and the Appearance/Reality Distinction." *Oxford Studies in Philosophy of Mind* 4: 380–413.

Loar, Brian. 1981. *Mind and Meaning*. Cambridge University Press.

Loar, Brian. 1990. "Phenomenal States." In *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, edited by James Tomberlin. Ridgeview.

Loar, Brian. 1997. "Phenomenal States (Revised Version)." In *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere. MIT Press.

Locke, John. [1689] 1975. *An Essay Concerning Human Understanding*. Clarendon Edition. Oxford University Press.

Long, Robert. 2023. "Introspective Capabilities in Large Language Models." *Journal of Consciousness Studies* 30 (9): 143–53. https://doi.org/10.53765/20512201.30.9.143.

Low, Philip, Jaak Panksepp, Diana Reiss, David Edelman, Bruno Van Swinderen, and Christof Koch. 2012. "The Cambridge Declaration on Consciousness." https://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf.

Lurz, Robert W. 2011. *Mindreading Animals: The Debate over What Animals Know about Other Minds*. A Bradford Book. MIT Press.
Lutz, Matt. 2014. "The 'Now What' Problem for Error Theory." *Philosophical Studies* 171 (2): 351–71. https://doi.org/10.1007/s11098-013-0275-7.
Lycan, William. 1996. *Consciousness and Experience*. MIT Press.
Lycan, William G. 2001. "Moore against the New Skeptics." *Philosophical Studies* 103 (1): 35–53. https://doi.org/10.1023/A:1010328721653.
Lyon, Pamela, and Franz Kuchling. 2021. "Valuing What Happens: A Biogenic Approach to Valence and (Potentially) Affect." *Philosophical Transactions of the Royal Society B* 376 (1820). https://doi.org/10.1098/rstb.2019.0752.
Lyons, Jack C. 2009. *Perception and Basic Beliefs: Zombies, Modules, and the Problem of the External World*. Oxford University Press.
Lyons, William. 1986. "The Disappearance of Introspection." *Philosophy of Science* 55 (4): 653–54.
Mackie, John. 1977. *Ethics: Inventing Right and Wrong*. Penguin.
Macphail, Euan. 1987. "The Comparative Psychology of Intelligence." *Behavioral and Brain Sciences* 10: 645–95.
Mandik, P. 2016. "Meta-Illusionism and Qualia Quietism." *Journal of Consciousness Studies* 23 (11–12): 140–48.
Maniscalco, Brian, and Hakwan Lau. 2012. "A Signal Detection Theoretic Approach for Estimating Metacognitive Sensitivity from Confidence Ratings." *Consciousness and Cognition* 21 (1): 422–30. https://doi.org/10.1016/j.concog.2011.09.021.
Marcotte, Roxanne. 2006. "L'aperception de Soi Chez Shihab al-Din al-Suhrawardi et l'héritage Avicennien." *Laval Théologique et Philosophique* 62 (3): 529–51.
Marcotte, Roxanne. 2012. "Suhrawardi." In *The Stanford Encyclopedia of Philosophy*, Summer 2012, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. http://plato.stanford.edu/archives/sum2012/entries/suhrawardi/.
Margolis, Eric, and Stephen Laurence. 2023. "Concepts." In *The Stanford Encyclopedia of Philosophy*, Fall 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entries/concepts/.
Margulis, Lynn. 2001. "The Conscious Cell." *Annals of the New York Academy of Sciences* 929 (1): 55–70. https://doi.org/10.1111/j.1749-6632.2001.tb05707.x.
Martínez, Manolo. 2013. "Teleosemantics and Indeterminacy." *Dialectica* 67 (4): 427–53. https://doi.org/10.1111/1746-8361.12039.
Mashour, George A., and Michael T. Alkire. 2013. "Evolution of Consciousness: Phylogeny, Ontogeny, and Emergence from General Anesthesia." *Proceedings of the National Academy of Sciences* 110 (supplement 2): 10357–64. https://doi.org/10.1073/pnas.1301188110.
Mashour, George A., Pieter Roelfsema, Jean-Pierre Changeux, and Stanislas Dehaene. 2020. "Conscious Processing and the Global Neuronal Workspace Hypothesis." *Neuron* 105 (5): 776–98. https://doi.org/10.1016/j.neuron.2020.01.026.
Matson, Wallace. 1966. "Why Isn't the Mind-Body Problem Ancient?" In *Mind, Matter, and Method*, NED-New edition, edited by Paul Feyerabend and Grover Maxwell. Essays in Philosophy and Science in Honor of Herbert Feigl. University of Minnesota Press.
Maund, Barry. 2022. "Color." In *The Stanford Encyclopedia of Philosophy*, Spring 2022, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2022/entries/color/.
Maxwell, Grover. 1979. "Rigid Designators and Mind-Brain Identity." *Minnesota Studies in the Philosophy of Science* 9: 365–403.
McCloskey, Michael. 1983. "Naive Theories of Motion." In *Mental Models*, edited by Derdre Gentner and Albert Stevens. Lawrence Erlbaum Associates.
McGinn, Colin. 1989. "Can We Solve the Mind-Body Problem?" *Mind* 98 (July): 349–66.

McGinn, Colin. 1993. "Consciousness and Cosmology: Hyperdualism Ventilated." In *Consciousness. Psychological and Philosophical Essays*, edited by Martin Davies and Glyn W. Humphreys. Blackwell.

McGinn, Colin. 2011. "The Paradox of Analysis." In *Truth by Analysis: Games, Names, and Philosophy*, edited by Colin McGinn. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199856145.003.0004.

Mckilliam, Andy. 2025. "Natural Kind Reasoning in Consciousness Science: An Alternative to Theory Testing." *Noûs* 59 (3): 634–51. https://doi.org/10.1111/nous.12526.

McQueen, K. J. 2019. "Illusionist Integrated Information Theory." *Journal of Consciousness Studies* 26 (5–6): 141–69.

Mellor, David J. 2019. "Welfare-Aligned Sentience: Enhanced Capacities to Experience, Interact, Anticipate, Choose and Survive." *Animals* 9 (7): 440. https://doi.org/10.3390/ani9070440.

Melnyk, Andrew. 2003. "Papineau on the Intuition of Distinctness." *SWIF Philosophy of Mind* 4. https://philarchive.org/archive/MELPOT.

Mendelovici, Angela. 2013. "Reliable Misrepresentation and Tracking Theories of Mental Representation." *Philosophical Studies* 165 (2): 421–43. https://doi.org/10.1007/s11098-012-9966-8.

Merleau-Ponty, Maurice. [1964] 1968. *The Visible and the Invisible*. Translated by Alphonso Lingis. Northwestern University Press.

Merlo, Giovanni. 2020. "Appearance, Reality and the Meta-Problem of Consciousness." *Journal of Consciousness Studies* 27 (5–6): 120–30.

Metzinger, Thomas. 2021. "Artificial Suffering: An Argument for a Global Moratorium on Synthetic Phenomenology." *Journal of Artificial Intelligence and Consciousness* 08 (01): 43–66. https://doi.org/10.1142/S270507852150003X.

Michaelian, Kourken. 2011. "Is Memory a Natural Kind?" *Memory Studies* 4 (2): 170–89. https://doi.org/10.1177/1750698010374287.

Michel, Matthias. 2019. "Consciousness Science Underdetermined: A Short History of Endless Debates." *Ergo, an Open Access Journal of Philosophy* 6 (20201214). https://doi.org/10.3998/ergo.12405314.0006.028.

Michel, Matthias. 2023. "Confidence in Consciousness Research." *WIREs Cognitive Science* 14 (2): e1628. https://doi.org/10.1002/wcs.1628.

Michel, Matthias. 2026. "The Perceptual Reality Monitoring Theory." In *Scientific Theories of Consciousness: The Grand Tour*, edited by Michael Herzog, Aaron Schurger, and Adrien Doerig. Cambridge University Press.

Miłkowski, Marcin. 2013. "Reverse-Engineering in Cognitive-Science." In *Regarding Mind, Naturally*, edited by Marcin Miłkowski and Konrad Talmont-Kaminski. Cambridge Scholars Press.

Millikan, Ruth. 1984. *Language, Thought and Other Biological Categories*. MIT Press.

Montague, Michelle. 2017. "Perception and Cognitive Phenomenology." *Philosophical Studies* 174 (8): 2045–62. JSTOR. https://doi.org/doi.org/10.1007/s11098-016-0787-z.

Montero, Barbara, and David Papineau. 2005. "A Defense of the Via Negativa Argument for Physicalism." *Analysis* 65 (3): 233–37.

Moore, Andrew. 2013. "Hedonism." In *The Stanford Encyclopedia of Philosophy (Winter 2013 Edition)*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2013/entries/hedonism.

Moore, Georges Edward. 1939. "Proof of an External World." *Proceedings of the British Academy* 25: 273–300.

Mørch, Hedda Hassel. 2017. "The Evolutionary Argument for Phenomenal Powers." *Philosophical Perspectives* 31 (1): 293–316. https://doi.org/10.1111/phpe.12096.

Mørch, Hedda Hassel. 2020. "The Phenomenal Powers View and the Meta-Problem of Consciousness." *Journal of Consciousness Studies* 27 (5–6): 131–42.

Muehlhauser, Luke. 2017. "Report on Consciousness and Moral Patienthood." *Open Philanthropy*. https://www.openphilanthropy.org/2017-report-consciousness-and-moral-patienthood.

Murdoch, Iris. 1980. *The Sovereignty of Good*. Repr. of the 1970 ed. Routledge & Kegan Paul.

Murez, Michael, and Enzo Thouzeau-Corpel. 2024. "Unbelievable yet True: Physical-Phenomenal Identities and the Mental Files Model." *Revue de Métaphysique et de Morale* 121 (1): 9–27.

Murray, Dylan, and Eddy Nahmias. 2014. "Explaining Away Incompatibilist Intuitions." *Philosophy and Phenomenological Research* 88 (2): 434–67. https://doi.org/10.1111/j.1933-1592.2012.00609.x.

Nadelhoffer, Thomas, David Rose, Wesley Buckwalter, and Shaun Nichols. 2020. "Natural Compatibilism, Indeterminism, and Intrusive Metaphysics." *Cognitive Science* 44 (8): e12873. https://doi.org/10.1111/cogs.12873.

Nagel, Thomas. 1974. "What Is It like to Be a Bat?" *Philosophical Review* 83 (October): 435–50.

Nagel, Thomas. 1979. *Mortal Questions*. Cambridge University Press.

Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. 2005. "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility." *Philosophical Psychology* 18 (5): 561–84. https://doi.org/10.1080/09515080500264180.

Neander, Karen. 2017. *A Mark of the Mental: In Defense of Informational Teleosemantics*. Life and Mind. MIT Press.

Nelkin, Norton. 1989. "Propositional Attitudes and Consciousness." *Philosophy and Phenomenological Research* 49 (3): 413–30.

Nichols, Shaun, and Joshua Knobe. 2007. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." *Noûs* 41 (4): 663–85. https://doi.org/10.1111/j.1468-0068.2007.00666.x.

Nichols, Shaun, and Stephen Stich. 2003. "How to Read Your Own Mind: A Cognitive Theory of Self-Consciousness." In *Consciousness: New Philosophical Perspectives*, edited by Quentin Smith and Aleksandar Jokic. Oxford University Press.

Nida-Rümelin, Martine. 2007. "Grasping Phenomenal Properties." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter and Sven Walter. Oxford University Press.

Nida-Rümelin, Martine. 2016. "The Illusion of Illusionism." *Journal of Consciousness Studies* 23 (11–12): 160–71.

Nida-Rümelin, Martine, and Donnchadh O'Conaill. 2021. "Qualia: The Knowledge Argument." In *The Stanford Encyclopedia of Philosophy*, Summer 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/qualia-knowledge/.

Nietzsche, Friedrich. [1883] 1954a. *Thus Spoke Zarathustra*. Translated by Walter Kaufmann. Viking.

Nietzsche, Friedrich. [1888] 1954b. *Twilights of the Idols*. Translated by Walter Kaufmann. Viking.

Nietzsche, Friedrich. [1882] 1974. *The Gay Science: With a Prelude in Rhymes and an Appendix of Songs*. 1st ed. Translated by Walter Kaufmann. Vintage.

Niikawa, Takuya. 2021. "Illusionism and Definitions of Phenomenal Consciousness." *Philosophical Studies* 178 (1): 1–21. https://doi.org/10.1007/s11098-020-01418-x.

Niikawa, Takuya, and Uriah Kriegel. 2025. "The Sublime of Consciousness." *British Journal of Aesthetics* 65 (1): 113–30. https://doi.org/10.1093/aesthj/ayae005.

Nisbett, Richard, and Timothy Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (3): 231–59.

Nussbaum, Martha C. 2024. *Justice for Animals: Our Collective Responsibility*. Reprint edition. Simon & Schuster.

O'Keefe, Tim. 2016. "Hedonistic Theories of Well-Being in Antiquity." In *The Routledge Handbook of Philosophy of Well-Being*, edited by Guy Fletcher. Routledge. https://doi.org/10.4324/9781315682266.ch3.

Olsson, Erik. 2023. "Coherentist Theories of Epistemic Justification." In *The Stanford Encyclopedia of Philosophy*, Winter 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/justep-coherence/.

Orijemie, Emuobosa Akpo. 2021. "Cultural Practices and Sustainable Management of Wetlands in Nigeria." *Proceedings of the 5th International Young Earth Scientists (YES) Congress*, 58–61. https://doi.org/10.2312/YES19.10.

Overgaard, Morten, Mika Koivisto, Thomas Alrik Sørensen, Signe Vangkilde, and Antti Revonsuo. 2006. "The Electrophysiology of Introspection." *Consciousness and Cognition* 15 (4): 662–72. https://doi.org/10.1016/j.concog.2006.05.002.

Papineau, David. 1984. "Representation and Explanation." *Philosophy of Science* 51 (December): 550–72.

Papineau, David. 1993a. *Philosophical Naturalism*. Blackwell.

Papineau, David. 1993b. "Physicalism, Consciousness, and the Antipathetic Fallacy." *Australasian Journal of Philosophy* 71: 169–83.

Papineau, David. 2002. *Thinking about Consciousness*. Oxford University Press.

Papineau, David. 2007. "Phenomenal and Perceptual Concepts." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter and Sven Walter. New York: Oxford University Press.

Papineau, David. 2008. "Explanatory Gaps and Dualist Intuitions." In *Frontiers of Consciousness*, edited by Lawrence Weiskrantz and Martin Davies. Oxford University Press.

Papineau, David. 2011. "What Exactly Is the Explanatory Gap?" *Philosophia* 39 (1): 5–19. https://doi.org/10.1007/s11406-010-9273-6.

Pappas, George. 2014. "Internalist vs. Externalist Conceptions of Epistemic Justification." In *The Stanford Encyclopedia of Philosophy*, Spring 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/justep-intext/.

Parfit, Derek. 1987. *Reasons and Persons*. 1. issued in paperback (with corr.), Reprinted with further corr. Clarendon Press.

Parfit, Derek. 2011. *On What Matters*. Vol. 1. Oxford University Press.

Patterson, Stephen J., James M. Robinson, and Hans-Gebhard Bethge. 2011. *The Fifth Gospel: The Gospel of Thomas Comes of Age*. New ed. T & t Clark Ltd.

Peacocke, Christopher. 1998a. "Implicit Conceptions, the 'A Priori,' and the Identity of Concepts." *Philosophical Issues* 9: 121. https://doi.org/10.2307/1522964.

Peacocke, Christopher. 1998b. "Implicit Conceptions, Understanding and Rationality." *Philosophical Issues* 9: 43. https://doi.org/10.2307/1522959.

Pereboom, Derk. 2009. "Consciousness and Introspective Inaccuracy." In *Appearance, Reality, and the Good: Themes from the Philosophy of Robert M. Adams*, edited by Larry Jorgensen and Samuel Newlands. Oxford University Press.

Pereboom, Derk. 2011. *Consciousness and the Prospects of Physicalism*. Oxford University Press.

Pereboom, Derk. 2019. "Russellian Monism, Introspective Inaccuracy, and the Illusion Meta-Problem of Consciousness." *Journal of Consciousness Studies* 26 (9–10): 182–93.

Peressini, Anthony. 2014. "Blurring Two Conceptions of Subjective Experience: Folk versus Philosophical Phenomenality." *Philosophical Psychology* 27 (6): 862–89. https://doi.org/10.1080/09515089.2013.793150.

Perez, Ethan, and Robert Long. 2023. "Towards Evaluating AI Systems for Moral Status Using Self-Reports." arXiv.Org, November 14. https://arxiv.org/abs/2311.08576v1.

Petitmengin, Claire. 2006. "Describing One's Subjective Experience in the Second Person: An Interview Method for the Science of Consciousness." *Phenomenology and the Cognitive Sciences* 5 (3): 229–69. https://doi.org/10.1007/s11097-006-9022-2.

Phillips, Ian. 2011. "Perception and Iconic Memory: What Sperling Doesn't Show." *Mind and Language* 26 (4): 381–411. https://doi.org/10.1111/j.1468-0017.2011.01422.x.

Pitt, David. 2004. "The Phenomenology of Cognition or What Is It like to Think That P?" *Philosophy and Phenomenological Research* 69: 1–36. https://doi.org/10.1111/j.1933-1592.2004.tb00382.x.

Prinz, Jesse J. 2003. "Level-Headed Mysterianism and Artificial Experience." *Journal of Consciousness Studies* 10 (4–5): 111–32.

Pritchard, Duncan. 2012. *Epistemological Disjunctivism*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199557912.001.0001.

Pryor, James. 2000. "The Skeptic and the Dogmatist." *Noûs* 34 (4): 517–49. https://doi.org/10.1111/0029-4624.00277.

Pryor, James. 2004. "What's Wrong with Moore's Argument?" *Philosophical Issues* 14 (1): 349–78. https://doi.org/10.1111/j.1533-6077.2004.00034.x.

Pylyshyn, Zenon. 1999. "Is Vision Continuous with Cognition?: The Case for Cognitive Impenetrability of Visual Perception." *Behavioral and Brain Sciences* 22 (3): 341–65. https://doi.org/10.1017/S0140525X99002022.

Quine, Willard Van Orman. 1960. *Word and Object*. MIT Press.

Rawls, John. 1971. *A Theory of Justice*. 1st ed. Harvard University Press.

Regan, Tom. 2004. *The Case for Animal Rights*. University of California Press.

Rey, Georges. 1983. "A Reason for Doubting the Existence of Consciousness." In *Consciousness and Self-Regulation: Volume 3: Advances in Research and Theory*, edited by Richard J. Davidson, Gary E. Schwartz, and David Shapiro. Springer US. https://doi.org/10.1007/978-1-4615-9317-1_1.

Rey, Georges. 1993. "The Unavailability of What We Mean." *Grazer Philosophische Studien* 46: 61–101. https://doi.org/10.5840/gps1993464.

Rey, Georges. 1995. "Towards a Projectivist Account of Conscious Experience." In *Conscious Experience*, edited by Thomas Metzinger. Ferdinand Schoningh.

Rey, Georges. 1997. "A Question about Consciousness." In *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere. MIT Press.

Rey, Georges. 1998. "A Naturalistic A Priori." *Philosophical Studies* 92 (1): 25–43. https://doi.org/10.1023/A:1017155400164.

Rey, Georges. 2007. "Phenomenal Content and the Richness and Determinacy of Colour Experience." *Journal of Consciousness Studies* 14 (9): 112–31.

Rice, Christopher M. 2013. "Defending the Objective List Theory of Well-Being." *Ratio* 26 (2): 196–211. https://doi.org/10.1111/rati.12007.

Rinard, Susanna. 2013. "Why Philosophy Can Overturn Common Sense1." In *Oxford Studies in Epistemology Volume 4*, edited by Tamar Szabó Gendler and John Hawthorne. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672707.003.0007.

Robinson, William S. 2022. "Dis-Illusioning Experiences." *Review of Philosophy and Psychology* 14: 1219–36. https://doi.org/10.1007/s13164-022-00630-4.

Roelofs, Luke. 2019. *Combining Minds: How to Think about Composite Subjectivity*. Oxford University Press.

Roelofs, Luke. 2023. "Sentientism, Motivation, and Philosophical Vulcans." *Pacific Philosophical Quarterly* 104 (2): 301–23. https://doi.org/10.1111/papq.12420.

Rollin, Bernard E. 1993. "Animal Welfare, Science, and Value." *Journal of Agricultural and Environmental Ethics* 1993 (6): 44–50.

Rollin, Bernard E. 2011. "Animal Pain: What It Is and Why It Matters." *Journal of Ethics* 15 (4): 425–37. https://doi.org/10.1007/s10892-010-9090-y.

Rorty, Richard. 1965. "Mind-Body Identity, Privacy and Categories." *Review of Metaphysics* 19: 24–54.

Rosenthal, David. 1993. "Thinking That One Thinks." In *Consciousness: Psychological and Philosophical Essays*, edited by Martin Davies and Glyn W. Humphreys. Readings in Mind and Language, Vol. 2. Blackwell Publishing.

Rosenthal, David. 1999. "Sensory Quality and the Relocation Story." *Philosophical Topics* 26 (1/2): 321–50. JSTOR. https://doi.org/10.5840/philtopics1999261/227.

Rosenthal, David. 2005. *Consciousness and Mind*. Oxford University Press.
Rosenthal, David. 2010. "How to Think about Mental Qualities." *Philosophical Issues* 20 (1): 368–93. https://doi.org/10.1111/j.1533-6077.2010.00190.x.
Rosenthal, David. 2019a. "Chalmers' Meta-Problem." *Journal of Consciousness Studies* 26 (9–10): 194–204.
Rosenthal, David. 2019b. "Consciousness and Confidence." *Neuropsychologia* 128 (May): 255–65. https://doi.org/10.1016/j.neuropsychologia.2018.01.018.
Rosenthal, David. 2022. "Mental Appearance and Mental Reality." In *Qualitative Consciousness: Themes from the Philosophy of David Rosenthal*, edited by Josh Weisberg. Cambridge University Press. Cambridge Core. https://doi.org/10.1017/9781108768085.018.
Ross, Alf. 1941. "On the Illusion of Consciousness: Do Mental Data ('Phenomena of Consciousness') Exist?" *Theoria* 7 (3): 3. https://doi.org/10.1111/j.1755-2567.1941.tb01039.x.
Rossi, Mauro, and Christine Tappolet. 2022. "Well-Being as Fitting Happiness." In *Fittingness*, edited by Christopher Howard and Rach Cosker-Rowland. Oxford University Press.
Rupert, Robert D. 2013. "Memory, Natural Kinds, and Cognitive Extension; or, Martians Don't Remember, and Cognitive Science Is Not about Cognition." *Review of Philosophy and Psychology* 4 (1): 25–47. https://doi.org/10.1007/s13164-012-0129-9.
Russell, Bertrand. 1911. "Knowledge by Acquaintance and Knowledge by Description." *Proceedings of the Aristotelian Society* 11 (1): 108–28. https://doi.org/10.1093/aristotelian/11.1.108.
Russell, Bertrand. 1912. *The Problems of Philosophy*. Oxford University Press.
Russell, Bertrand. 1950. "Introduction: Materialism, Past and Present." In *The History of Materialism*, by Friedrich Albert Lange. The Humanities Press.
Russell, Bruce. 2020. "A Priori Justification and Knowledge." In *The Stanford Encyclopedia of Philosophy*, Summer 2024, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2024/entries/apriori/.
Ruyer, Raymond. 1930. *Esquisse d'une Philosophie de La Structure*. Félix Alcan.
Ruyer, Raymond. [1937] 2020. *La Conscience et Le Corps*. Vrin.
Saad, Bradford, and Adam Bradley. 2022. "Digital Suffering: Why It's a Problem and How to Prevent It." *Inquiry* 68 (7): 2110–45. https://doi.org/10.1080/0020174X.2022.2144442.
Saigusa, Tetsu, Atsushi Tero, Toshiyuki Nakagaki, and Yoshiki Kuramoto. 2008. "Amoebae Anticipate Periodic Events." *Physical Review Letters* 100 (1): 018101. https://doi.org/10.1103/PhysRevLett.100.018101.
Schlicht, Tobias. 2018. "A Methodological Dilemma for Investigating Consciousness Empirically." *Consciousness and Cognition* 66: 91–100. https://doi.org/10.1016/j.concog.2018.11.002.
Schlicht, Tobias, and Krzysztof Dolega. 2021. "You Can't Always Get What You Want: Predictive Processing and Consciousness." *Philosophy and the Mind Sciences* 2. https://doi.org/10.33735/phimisci.2021.80.
Schmidt, Marco F.H., and Hannes Rakoczy. 2023. "Children's Acquisition and Application of Norms." *Annual Review of Developmental Psychology* 5 (1): 193–215. https://doi.org/10.1146/annurev-devpsych-120621-034731.
Schrödinger, Erwin. 1958. *Mind and Matter*. Cambridge University Press.
Schwarz, Wolfgang. 2018. "Imaginary Foundations." *Ergo* 5 (29): 764–89. https://doi.org/10.3917/rmm.241.0009.
Schwarz, Wolfgang. 2019. "From Sensor Variables to Phenomenal Facts." *Journal of Consciousness Studies* 26 (9–10): 217–27.
Schwitzgebel, Eric. 2011. *Perplexities of Consciousness*. MIT Press.
Schwitzgebel, Eric. 2012. "Introspection, What?" In *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744794.003.0001.

Schwitzgebel, Eric. 2015. "If Materialism Is True, the United States Is Probably Conscious." *Philosophical Studies* 172 (7): 1697–721. https://doi.org/10.1007/s11098-014-0387-8.

Schwitzgebel, Eric. 2016a. "Is the United States Phenomenally Conscious? Reply to Kammerer." *Philosophia* 44 (3): 877–83. https://doi.org/10.1007/s11406-016-9725-8.

Schwitzgebel, Eric. 2016b. "Phenomenal Consciousness, Defined and Defended as Innocently as I Can Manage." *Journal of Consciousness Studies* 23 (11–12): 224–35.

Schwitzgebel, Eric. 2023. "Borderline Consciousness, When It's Neither Determinately True nor Determinately False That Experience Is Present." *Philosophical Studies* 180 (3415–39). https://doi.org/10.1007/s11098-023-02042-1.

Schwitzgebel, Eric. 2025. "Kings, Wizards, and Illusionism about Consciousness." *The Splintered Mind*, March 6. https://schwitzsplinters.blogspot.com/2025/03/kings-wizards-and-illusionism-about.html.

Schwitzgebel, Eric, and Mara Garza. 2015. "A Defense of the Rights of Artificial Intelligences: Defense of the Rights of Artificial Intelligences." *Midwest Studies In Philosophy* 39 (1): 98–119. https://doi.org/10.1111/misp.12032.

Scribner, Vaughn. 2017. "'Such Monsters Do Exist in Nature': Mermaids, Tritons, and the Science of Wonder in Eighteenth-Century Europe." *Itinerario* 41 (3): 507–38. https://doi.org/10.1017/S0165115317000663.

Seager, William. 2001. "Consciousness, Value and Functionalism." *Psyche* 7: 1–19.

Searle, John. 1992. *The Rediscovery of the Mind*. MIT Press.

Searle, John. 1997. *The Mystery of Consciousness*. The New York Review of Books.

Sebo, Jeff. 2017. "Agency and Moral Status." *Journal of Moral Philosophy* 14 (1): 1–22. https://doi.org/10.1163/17455243-46810046.

Sebo, Jeff. 2018. "The Moral Problem of Other Minds." *The Harvard Review of Philosophy* 25: 51–70. https://doi.org/10.5840/harvardreview20185913.

Segundo-Ortin, Miguel, and Paco Calvo. 2022. "Consciousness and Cognition in Plants." *WIREs Cognitive Science* 13 (2): e1578. https://doi.org/10.1002/wcs.1578.

Seth, Anil K. 2025. "Conscious Artificial Intelligence and Biological Naturalism." *Behavioral and Brain Sciences*: 1–4210. https://doi.org/1017/S0140525X25000032.

Seth, Anil K., Bernard J. Baars, and David B. Edelman. 2005. "Criteria for Consciousness in Humans and Other Mammals." *Consciousness and Cognition* 14 (1): 119–39. https://doi.org/10.1016/j.concog.2004.08.006.

Seth, Anil K., and Tim Bayne. 2022. "Theories of Consciousness." *Nature Reviews Neuroscience* 23 (7): 439–52. https://doi.org/10.1038/s41583-022-00587-4.

Shabasson, Daniel. 2022. "Illusionism about Phenomenal Consciousness: Explaining the Illusion." *Review of Philosophy and Psychology* 13: 427–53. https://doi.org/10.1007/s13164-021-00537-6.

Shafer-Landau, Russ. 1994. "Ethical Disagreement, Ethical Objectivism and Moral Indeterminacy." *Philosophy and Phenomenological Research* 54 (2): 331–44.

Shaffer, Jerome A. 1968. *Philosophy of Mind*. Prentice-Hall.

Shea, Nicholas. 2012. "Methodological Encounters with the Phenomenal Kind." *Philosophy and Phenomenological Research* 84 (2): 307–44. https://doi.org/10.1111/j.1933-1592.2010.00483.x.

Shepherd, Joshua. 2016. "Moral Conflict in the Minimally Conscious State." In *Finding Consciousness: The Neuroscience, Ethics, and Law of Severe Brain Damage*, edited by Walter Sinnott-Armstrong. Oxford University Press.

Shepherd, Joshua. 2018. *Consciousness and Moral Status*. Routledge Focus on Philosophy. Routledge, Taylor & Francis Group.

Shepherd, Joshua. 2022. "Non-Human Moral Status: Problems with Phenomenal Consciousness." *AJOB Neuroscience* 14 (2): 148–57. https://doi.org/10.1080/21507740.2022.2148770.

Shevlin, Henry. 2020. "Which Animals Matter? Comparing Approaches to Psychological Moral Status in Nonhuman Systems." *Philosophical Topics* 48 (1): 177–200. https://doi.org/10.5840/philtopics20204819.

Shevlin, Henry. 2021. "Non-human Consciousness and the Specificity Problem: A Modest Theoretical Proposal." *Mind & Language* 36 (2): 297–314. https://doi.org/10.1111/mila.12338.

Shields, Christopher. 2020a. "Aristotle's Psychology." In *The Stanford Encyclopedia of Philosophy*, Winter 2020, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/aristotle-psychology/.

Shields, Christopher. 2020b. "Controversies Surrounding Aristotle's Theory of Perception (Supplement to Aristotle's Psychology)." In *The Stanford Encyclopedia of Philosophy*, Winter 2020, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/aristotle-psychology/suppl3.html.

Shoemaker, Sydney. 1975. "Functionalism and Qualia." *Philosophical Studies* 27 (May): 291–315.

Shoemaker, Sydney. 1990. "Qualities and Qualia: What's in the Mind?" *Philosophy and Phenomenological Research Supplement* 50 (Supplement): 109–31.

Shoemaker, Sydney. 1994. "The Phenomenal Character of Experience." *Philosophy and Phenomenological Research* 54 (2): 291–314.

Shoemaker, Sydney. 2003. "Content, Character and Color." *Philosophical Issues* 13 (1): 253–78. https://doi.org/10.1111/1533-6077.00014.

Sider, Theodore. 2011. *Writing the Book of the World*. Oxford University Press. https://academic.oup.com/book/3171.

Sidgwick, Henry. 1907. *The Method of Ethics*. Thoemmes Press.

Siewert, Charles. 1998. *The Significance of Consciousness*. Princeton University Press.

Siewert, Charles. 2012. "On the Phenomenology of Introspection." In *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar. Oxford University Press.

Siewert, Charles. 2014. "Speaking Up for Consciousness." In *Current Controversies in Philosophy of Mind*, edited by Uriah Kriegel. Routledge.

Silins, Nicholas. 2024. "Perceptual Experience and Perceptual Justification." In *The Stanford Encyclopedia of Philosophy*, Fall 2024, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2024/entries/perception-justification/.

Simmons, Alison. 2015. "Perception in Early Modern Philosophy." In *The Oxford Handbook of Philosophy of Perception*, edited by Mohan Matthen. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199600472.013.015.

Simon, Jonathan. 2017. "Vagueness and Zombies: Why 'Phenomenally Conscious' Has No Borderline Cases." *Philosophical Studies* 174 (8): 2105–23. https://doi.org/10.1007/s11098-016-0790-4.

Simon, Jonathan. 2020. *Review of "Human and Animal Minds: The Consciousness Questions Laid to Rest,"* by Peter Carruthers. December 2. https://ndpr.nd.edu/reviews/human-and-animal-minds-the-consciousness-questions-laid-to-rest/.

Sims, Matthew, and Julian Kiverstein. 2022. "Externalized Memory in Slime Mould and the Extended (Non-Neuronal) Mind." *Cognitive Systems Research* 73: 26–35. https://doi.org/10.1016/j.cogsys.2021.12.001.

Singer, Peter. 1993. *Practical Ethics*. Cambridge University Press.

Smithies, Declan. 2013. "On the Unreliability of Introspection." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 165 (3): 1177–86. https://doi.org/10.1007/s11098-013-0150-6.

Smithies, Declan. 2019. *The Epistemic Role of Consciousness*. Oxford University Press.

Sneddon, Lynne U. 2019. "Evolution of Nociception and Pain: Evidence from Fish Models." *Philosophical Transactions of the Royal Society B: Biological Sciences* 374 (1785): 20190290. https://doi.org/10.1098/rstb.2019.0290.

Sosa, Ernest. 1991. *Knowledge in Perspective: Selected Essays in Epistemology*. 1st ed. Cambridge University Press. https://www.cambridge.org/core/product/identifier/9780511625299/type/book.

Spelke, Elizabeth S. 2000. "Core Knowledge." *American Psychologist* 55 (11): 1233–43. https://doi.org/10.1037/0003-066X.55.11.1233.

Spelke, Elizabeth S., and Katherine D. Kinzler. 2007. "Core Knowledge." *Developmental Science* 10 (1): 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x.

Spener, Maja. 2024. *Introspection: First-Person Access in Science and Agency*. Oxford University Press.

Sprigge, Timothy. 1988. *The Rational Foundations of Ethics*. Routledge.

Spurrett, David, and David Papineau. 1999. "A Note on the Completeness of 'Physics.'" *Analysis* 59 (1): 25–29.

Ståhl, Tomas. 2021. "The Amoral Atheist? A Cross-National Examination of Cultural, Motivational, and Cognitive Antecedents of Disbelief, and Their Implications for Morality." *PLOS ONE* 16 (2): e0246593. https://doi.org/10.1371/journal.pone.0246593.

Stampe, Dennis. 1977. "Towards a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy* 2 (1): 42–63.

Steup, Matthias, and Ram Neta. 2020. "Epistemology." In *The Stanford Encyclopedia of Philosophy*, Spring 2024, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2024/entries/epistemology/.

Stoljar, Daniel. 2005. "Physicalism and Phenomenal Concepts." *Mind and Language* 20 (2): 296–302. https://doi.org/10.1111/j.0268-1064.2005.00296.x.

Stoljar, Daniel. 2024. "Physicalism." In *The Stanford Encyclopedia of Philosophy*, Spring 2024, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2024/entries/physicalism.

Strawson, Galen. 1994. *Mental Reality*. MIT Press.

Strawson, Galen. 2003. "Real Materialism." In *Chomsky and His Critics*, 1st ed., edited by Louise M. Antony and Norbert Hornstein. Wiley. https://doi.org/10.1002/9780470690024.ch3.

Strawson, Galen. 2015. "Self-Intimation." *Phenomenology and the Cognitive Sciences* 14 (1): 1–31. https://doi.org/10.1007/s11097-013-9339-6.

Strawson, Galen. 2018. "The Consciousness Deniers." *New York Review of Books* March 13. https://www.nybooks.com/daily/2018/03/13/the-consciousness-deniers/.

Sumner, L.W. 1992. "Two Theories of the Good." *Social Philosophy and Policy* 9 (2): 1–14. Cambridge Core. https://doi.org/10.1017/S0265052500001370.

Sundström, Pär. 2008. "Is the Mystery an Illusion? Papineau on the Problem of Consciousness." *Synthese* 163 (2): 133–43. https://doi.org/10.1007/s11229-007-9193-6.

Swinburne, Richard. 2001. *Epistemic Justification*. Oxford University Press.

Synofzik, Matthis, Gottfried Vosgerau, and Albert Newen. 2008. "Beyond the Comparator Model: A Multifactorial Two-Step Account of Agency." *Consciousness and Cognition* 17 (1): 219–39. https://doi.org/10.1016/j.concog.2007.03.010.

Sytsma, Justin, and Edouard Machery. 2010. "Two Conceptions of Subjective Experience." *Philosophical Studies* 151 (2): 299–327. https://doi.org/10.1007/s11098-009-9439-x.

Sytsma, Justin, and Eyuphan Ozdemir. 2019. "No Problem: Evidence That the Concept of Phenomenal Consciousness Is Not Widespread." *Journal of Consciousness Studies* 26 (9–10): 241–56.

Tännsjö, Torbjörn. 1998. *Hedonistic Utilitarianism*. Edinburgh University Press.

Tartaglia, James. 2013. "Conceptualizing Physical Consciousness." *Philosophical Psychology* 26 (6): 817–38. https://doi.org/10.1080/09515089.2013.770940.

Tennant, Neil. 2007. "Mind, Mathematics and the Ignorabimusstreit." *British Journal for the History of Philosophy* 15 (4): 745–73. https://doi.org/10.1080/09608780701605036.

Thompson, Evan. 2022. "Could All Life Be Sentient?" *Journal of Consciousness Studies* 29 (3): 229–65. https://doi.org/10.53765/20512201.29.3.229.

Timmermans, Bert, and Axel Cleeremans. 2015. "How Can We Measure Awareness? An Overview of Current Methods." In *Behavioural Methods in Consciousness Research*, edited by Morten Overgaard, 21–46. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199688890.003.0003.

Titchener, Edward Bradford. 1901. *Experimental Psychology*. Macmillan.
Tononi, Giulio. 2012. "Integrated Information Theory of Consciousness: An Updated Account." *Archives Italiennes de Biologie* 150: 290–326. https://doi.org/10.4449/aib.v149i5.1388.
Tononi, Giulio, Melanie Boly, Marcello Massimini, and Christof Koch. 2016. "Integrated Information Theory: From Consciousness to Its Physical Substrate." *Nature Reviews Neuroscience* 17 (7): 450–61. https://doi.org/10.1038/nrn.2016.44.
Tsuchiya, Naotsugu, Melanie Wilke, Stefan Frässle, and Victor A.F. Lamme. 2015. "No-Report Paradigms: Extracting the True Neural Correlates of Consciousness." *Trends in Cognitive Sciences* 19 (12): 757–70. https://doi.org/10.1016/j.tics.2015.10.002.
Turgenev, Ivan Sergeevich. [1862] 2009. *Fathers and Sons*. Translated by Peter Carson. Penguin Classics. Penguin Books.
Turri, John. 2016. "A New Paradigm for Epistemology From Reliabilism to Abilism." *Ergo, an Open Access Journal of Philosophy* 3 (8): 189–231. https://doi.org/10.3998/ergo.12405314.0003.008.
Turri, John, Mark Alfano, and John Greco. 2021. "Virtue Epistemology." In *The Stanford Encyclopedia of Philosophy*, Winter 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/epistemology-virtue/.
Tversky, Amos, and Daniel Kahneman. 1983. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90: 293–315.
Tye, Michael. 1995. *Ten Problems of Consciousness : A Representational Theory of the Phenomenal Mind*. Representation and Mind. MIT Press.
Tye, Michael. 1999. "Phenomenal Consciousness : The Explanatory Gap as a Cognitive Illusion." *Mind* 108 (432): 705–25. https://doi.org/10.1093/mind/108.432.705.
Tye, Michael. 2002. "Visual Qualia and Visual Content Revisited." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David Chalmers. Oxford University Press.
Tye, Michael. 2003. "A Theory of Phenomenal Concepts." In *Minds and Persons*, edited by Anthony O'Hear. Cambridge University Press.
Tye, Michael. 2017. *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190278014.001.0001.
Tye, Michael. 2021. *Vagueness and the Evolution of Consciousness: Through the Looking Glass*. Oxford University Press. https://doi.org/10.1093/oso/9780198867234.001.0001.
Tyndall, John. 1871. *Fragments of Science for Unscientific People*. New York : D. Appleton.
Vahid, Hamid. 2024. "Doxastic Conservatism." In *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/doxastic-conservatism/.
Veit, Walter, and Heather Browning. 2023. "Defending Sentientism." *AJOB Neuroscience* 14 (2): 168–70. https://doi.org/10.1080/21507740.2023.2188292.
Voltaire. [1764] 1924. *The Philosophical Dictionary*. Translated by H.I. Woolf. Knopf.
Wall, Steven. 2021. "Perfectionism in Moral and Political Philosophy." In *The Stanford Encyclopedia of Philosophy*, Fall 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/perfectionism-moral/.
Watson, John B. 1913. "Psychology as the Behaviorist Views It." *Psychological Review* 101 (2): 2. https://doi.org/10.1037/0033-295X.101.2.248.
Watson, John B. 1924. *Psychology from the Standpoint of a Behaviorist*. 2nd ed. J.B. Lippincott Company.
Weil, Simone. [1947] 2003. *Gravity and Grace*. Edited by Gustave Thibon. Translated by Emma Crawford and Mario Von der Ruhr. Taylor & Francis e-Library.
Weisberg, Josh. 2023. *Explanatory Optimism about the Hard Problem of Consciousness*. Routledge. https://doi.org/10.4324/9781003411581.
Weiskrantz, Lawrence. 2007. "The Case of Blindsight." In *The Blackwell Companion to Consciousness*, edited by Max Velmans and Schneider. Blackwell Publishing.

Wilcox, Marc G. 2020. "Animals and the Agency Account of Moral Status." *Philosophical Studies* 177 (7): 1879–99. https://doi.org/10.1007/s11098-019-01289-x.

Williams, D.C. 1934. "Scientific Method and the Existence of Consciousness." *Psychological Review* 41 (5): 461–79. https://doi.org/10.1037/h0070811.

Woodard, Christopher. 2015. "Hybrid Theories." In *The Routledge Handbook of Philosophy of Well-Being*, edited by Guy Fletcher. Routledge.

Wu, Wayne. 2018. "The Neuroscience of Consciousness." In *The Stanford Encyclopedia of Philosophy*, Winter 2018, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2018/entries/consciousness-neuroscience/.

Wu, Wayne. 2023. *Movements of the Mind: A Theory of Attention, Intention and Action*. Oxford University Press. https://doi.org/10.1093/oso/9780192866899.001.0001.

Wundt, Wilhelm. 1874. *Grundzüge Der Physiologischen Psychologie*. Wilhelm Engelmann.

Yudkowsky, Eliezer. 2015. "Zombies! Zombies?" In *Rationality: From AI to Zombies*. Machine Intelligence Research Institute. https://www.readthesequences.com/Zombies-Zombies.

Zilio, Federico, and Andrea Lavazza. 2023. "Consciousness in a Rotor? Science and Ethics of Potentially Conscious Human Cerebral Organoids." *AJOB Neuroscience* 14 (2): 178–96. https://doi.org/10.1080/21507740.2023.2173329.

Index

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.