



**EXPLANATORY OPTIMISM
ABOUT THE HARD PROBLEM
OF CONSCIOUSNESS**

Josh Weisberg



ROUTLEDGE

Explanatory Optimism about the Hard Problem of Consciousness

Explanatory Optimism about the Hard Problem of Consciousness argues that despite the worries of explanatory pessimists, consciousness can be fully explained in “easy” scientific terms. The widespread intuition that consciousness poses a hard problem is plausibly based on how consciousness appears to us in first-person access. The book offers a debunking argument to undercut the justificatory link between the first-person appearances and our hard problem intuitions.

The key step in the debunking argument involves the development and defense of an empirical model of first-person access: Automated Compression Theory (ACT). ACT holds that first-person access to consciousness is accomplished by automated accessing of compressed sensory information. Because of the distorting nature of this compressed access, it seems to subjects that consciousness possesses “exceptional” properties—properties leading to the hard problem—even though no such properties are present. If there are no exceptional properties to explain, then an explanation in easy terms can fully account for conscious experience. The book presents a range of empirical evidence for ACT and concludes that the burden of proof is now on the pessimists to show why we shouldn’t be optimistic about explaining consciousness.

Josh Weisberg is Associate Professor of Philosophy at the University of Houston. He is the author of *Consciousness* (2014) and editor of *Qualitative Consciousness: Themes from the Philosophy of David Rosenthal* (2022) and has published a range of articles in philosophy of mind and consciousness studies.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Who can guess what would be the result of continuing the pursuit of science for ten thousand years, with the activity of the last hundred? And if it were to go on for a million, or a billion, or any number of years you please, how is it possible to say that there is any question which might not ultimately be solved?

—C.S. Peirce, “**How to Make Our Ideas Clear**”



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Explanatory Optimism about the Hard Problem of Consciousness

Josh Weisberg

First published 2024
by Routledge
605 Third Avenue, New York, NY 10158, USA

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN, UK

Routledge is an imprint of the Taylor & Francis Group, an informa business.

© 2024 Taylor & Francis

The right of Josh Weisberg to be identified as author of this work has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

ISBN: 978-1-032-53343-8 (hbk)

ISBN: 978-1-032-53344-5 (pbk)

ISBN: 978-1-003-41158-1 (ebk)

DOI: 10.4324/9781003411581

Typeset in Times New Roman
by SPi Technologies India Pvt Ltd (Straive)

Contents

<i>Acknowledgements</i>	<i>viii</i>
1 Explanatory Optimism	1
2 The Hard Problem and Qualitative Inaccuracy	22
3 The Appearances to be Explained	50
4 Present and Past Debunkers	74
5 Automated Compression Theory	92
6 ACT, Expertise, and First-Person Access	114
7 Optimistic ACT at Work	138
<i>References</i>	<i>158</i>
<i>Index</i>	<i>172</i>

Acknowledgements

In writing this book, I've benefitted greatly from support, encouragement, conversation, and criticism from too many people to name. I'll try to single out those who've helped most directly, and I sincerely apologize to those I've embarrassingly left out. Various parts of this work were presented over the years at the Tucson "Towards a Science of Consciousness" conference, the Association for the Scientific Study of Consciousness annual meeting, the annual meeting of the Southern Society for Philosophy and Psychology, the annual meeting of the Society for Philosophy and Psychology, the CUNY Cognitive Science Speaker Series, the Jerrold J. Katz Memorial Lecture, CUNY Graduate Center, and the Rice Summer Workshop on the Interface Between Perception and Thought. I offer my sincere thanks to audiences and commentators at these events. In addition, a precursor of some of the key ideas herein, notably the ACT (though labeled "CAT") was presented at Higher-Order Theories and Mental Qualities: A Conference in Honor of David Rosenthal (Taipei, Taiwan, August 2015). I give my sincere thanks to Tim Lane and the conference organizers and the generous support of the Taiwan Medical University and Academia Sinica. An incomplete list of those who've contributed to my thinking about these issues in various conversations over the years includes Kathleen Akins, Torin Alter, Tim Bayne, Dave Beisecker, Jake Berger, Jared Blank, Ned Block, Brit Brogaard, Richard Brown, Alex Byrne, Gregg Caruso, David Chalmers, Austen Clark, Sam Coleman, Daniel Dennett, Rocco Gennaro, Jim Hitt, Bob Kentridge, Amy Kind, Bernie Kobes, Hakwan Lao, Andrew Lee, Joseph Levine, Doug Meehan, Myrto Mylopoulos, Miguel Angel Sebastian, Pär Sundström, Robert Van Gulick, and Liz Vlahos. I offer special thanks to Pete Mandik and Ken Williford. And my thinking on all matters about consciousness has been deeply influenced by my friend and thesis advisor David Rosenthal, who has my gratitude, particularly for shaping my ideas about the appearance and reality of conscious mental states. And thanks to the generous work of Luis Oliveira and Uriah Kriegel, who read all or part of the manuscript. Their comments and observations made this work far better than it might have

been. Thanks also to my wonderful and supportive colleagues at the University of Houston. Finally, thanks to my family for sticking with me through all this, especially my mom, Judy Weisberg, and my incredible wife, Ashley Hope.

This book is dedicated to my sons, Winnie and Frankie. You made it harder to write this book, in the best possible way. Love you!



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Explanatory Optimism

It is widely held that the problem of explaining consciousness is hard. David Chalmers, who coined the label “the hard problem of consciousness,” tells us:

[c]onsciousness is the biggest mystery. It may be the last outstanding obstacle to our quest for a scientific understanding of the universe. ... We have good reason to believe that consciousness arises from physical systems such as brains, but we have little idea how it arises or why it exists at all. ... We do not just lack a detailed theory; we are entirely in the dark about how consciousness fits into the natural order.

(Chalmers 1996, xi)

In a similar spirit, Colin McGinn writes,

[w]e know that brains are the *de facto* causal basis of consciousness, but we have, it seems, no understanding whatever of how this can be so. It strikes us as miraculous, eerie, even faintly comic. Somehow, we feel, the water of the physical brain is turned into the wine of consciousness, but we draw a total blank on the nature of this conversion.

(McGinn 1989, 529)

This is nothing new. In the late 19th century, Thomas Huxley worried, “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 Djin when Aladdin rubbed his lamp” (Huxley 1866, 193). And, in 1759, Samuel Johnson paused from more literary labors to note, “Matter can differ from matter only in form, bulk, density, motion and direction of motion: to which of these, however varied or combined, can consciousness be annexed?” Hard, indeed.

According to Chalmers, even if we knew everything that current science could tell us about the brain, we still wouldn't know *why* that brain is conscious. Contemporary science is plausibly in the business of laying out the structural, dynamical, and functional properties of things—laying out how the thing's parts are arranged, how those parts interact and change over time, and the role that those parts play in the thing's successful functioning. But, Chalmers contends, knowing all about the structure, dynamics, and function of the brain solves only “easy problems”—problems amenable to this sort of structural-causal-functional explanation. Knowing why the brain is conscious seems to be a problem of different kind—it is hard. Indeed, according to conventional philosophical wisdom, to have any hope of solving the hard problem, we'll likely have to radically alter our scientific worldview, expanding our ontology to include exceptional “phenomenal properties,” properties of experience whose nature is inexplicable in structural, dynamical, and functional terms. The hard problem thus appears to mark a break point for current science. New things, undreamt of from the contemporary perspective, may be required to move forward.¹

In this book, I argue that, despite its hard façade, the problem of explaining consciousness is easy. Consciousness—the full-blown subjective, phenomenal, “what it's like” kind of consciousness of sensing and feeling—can be explained in terms of structural, dynamical, and functional properties. It can be fully explained—without remainder, without lingering open questions or explanatory gaps—just like rocks and rockets, ships and sealing wax, respiration and perspiration.

My view, when put so bluntly, is widely seen to be a non-starter, even by those who are sympathetic to the idea that consciousness is ultimately physical in nature. Chalmers, commenting on this sort of approach, writes “that it appears to deny the manifest” (2003a, 109). He continues,

[t]o flatly deny ... without argument that there is a hard problem of consciousness over and above the easy problems, would be to make a highly counterintuitive claim that begs the important questions.

(109–110)

Of course, I will endeavor to make arguments, sound ones if I can. But the claim is seen by Chalmers and many others as “highly counterintuitive.” Worse yet, the claim may amount to a denial of the existence of consciousness altogether, if one thinks consciousness is explanatorily exceptional and hard by definition. And that would be truly unhinged: a conscious being arguing that consciousness doesn't exist. In a characteristically pointed passage, Galen Strawson puts this kind of eliminativist reasoning into perspective:

This particular denial is the strangest thing that has ever happened in the whole history of human thought, not just the whole history of philosophy. It falls, unfortunately, to philosophy, not religion, to reveal the deepest woo-woo of the human mind. I find this grievous, but, next to this denial, every known religious belief is only a little less sensible than the belief that grass is green.

(Strawson 2006, 5–6)

The goal of this book is to show that a rejection of the hard problem neither denies the manifest nor commits the “deepest woo-woo of the human mind.” I’ll argue that it makes good sense to reject the hard problem, once we see that support for the hard problem is plausibly *undercut*. To undercut the hard problem, I will develop a debunking story, one that shows why it seems so reasonable and obvious to think there is a hard problem even though there isn’t one. And once we have the story in hand, we’ll see that though the hard problem isn’t categorically ruled out, it lacks appropriate motivation. And that shouldn’t be so surprising: a central lesson of the historical sweep of modern science is that, often, seemingly intractable explanatory problems are solvable when looked at in a fresh light. A more plausible response to the problem of consciousness than thinking we need a radically revised ontology is instead concluding that we’re making an *epistemic mistake*. Consciousness is not explanatorily intractable, it just seems to be. And this problematic appearance can be explained in ordinary structural, dynamical, and functional terms. Contrary to appearances, there is no hard problem of consciousness.

I’m not, of course, the first person to make this sort of case. I stand on the shoulders of a number of giants and I’ll call them out in due course.² Still, the case, I think, is worth making, both because of an upsurge in work on the details of debunking arguments in general and because of a rich debate on why there even *appears* to be a problem of consciousness.³ In addition, there’s been a recent wave of radical approaches to consciousness from both philosophers and scientists—views abandoning the idea that we might explain consciousness using the resources of contemporary science.⁴ I will argue that the motivation for these theories is rooted in a mistake about the nature of consciousness, a mistake brought on by the misleading appearance of consciousness itself. For many, this will not dissolve the impression of a hard problem, no doubt—the misleading appearances will remain as “manifest” as ever, even if my debunking story is correct. But it will at least generate the need for the supporters of the hard problem to respond in detail, to defend what seems to many of them to be too obvious to require defense. But this obviousness is not enough to make the case, if I’m correct—the obviousness will be present with or without the hard problem. Consciousness is easy, though it seems hard, perhaps obviously so. But appearances can be deceptive, even here.

1.1 Explanatory Pessimists vs. Explanatory Optimists

However, it may seem that we already have conclusive reasons for thinking that, whatever may be true about other natural phenomena, consciousness just cannot be explained in structural, dynamical, or functional terms. First-person access to consciousness, it is argued, delivers a decisive verdict on this question—there is something fundamentally different here, something beyond anything explicable in terms of the structure, dynamics, and functioning of the physical brain. What’s more, that’s okay. It doesn’t necessarily entail that consciousness is nonphysical or inexplicable in broadly naturalistic terms or that we can’t find a secure place for it in a science-friendly ontology. Further, it may seem that something deep and special will be lost if consciousness turns out to be nothing more than what “easy” explanation allows. What of our intrinsically subjective lives, or the rich reality of our phenomenal experience, or perhaps even our purposes and values? Why tilt at this particular windmill if, seemingly, the chances of success are so low and the costs of success so high? Finally, my approach may look like a rather grim form of reduction and elimination, a narrow scientific quest to trap all reality in quasi-mathematical strictures—a blinkered vision of explanation and reality, ruling out by fiat much of what’s philosophically important and interesting. Why stick stubbornly to such an impoverished view of things? Perhaps the hard problem simply shows that there’s more than is dreamt of in a constricted explanatory framework limited to structural, dynamical, and functional properties.⁵

But my view, on the contrary, is that if we accept the hard problem, we are giving in to a kind of “explanatory pessimism,” one where we give up prematurely on the possibility of an informative explanation. Of course, the central issue is whether or not we’re being premature here. In response to such pessimism, I will defend what I’ll call “explanatory optimism.” It is the main burden of this book to show that such optimism is warranted, even when it comes to phenomenal consciousness. In this section, I’ll spell out what I mean by “explanatory optimism,” and I’ll show how it gains support from and meshes effectively with the naturalistic thinking inspired by modern science. Then I’ll argue that this scientifically informed explanatory optimism supports the idea that we should pursue explanation in structural, dynamical, and functional terms, if such an explanation is a viable option. Finally, I’ll show why these considerations recommend a debunking, undercutting approach to the hard problem, one that removes the justificatory grounds for explanatory pessimism about consciousness.

Explanatory optimism can be characterized as follows:

Explanatory optimism is a positive attitude toward the possibility of an informative explanation.

An informative explanation is one that explains a target explanandum in terms of other, better understood phenomena.⁶ If a phenomenon is held to be a primitive in an explanatory system or if it is held to be “strongly emergent” in an explanatory system, then it cannot be informatively explained. Something is held to be primitive in an explanatory system when it is posited without further informative explanation. Something is held to be strongly emergent when it is posited to be necessarily determined by other features described in an explanatory system but without the possibility of any informative explanation of how and why the phenomenon emerges. My claim is that explanatory optimism is warranted in the case of consciousness: we should be optimistic about the possibility of explaining consciousness in other, better understood terms—terms which do not involve ‘consciousness’ or its cognates and terms which aren’t themselves primitive or mysterious.

Why, then, should we be explanatory optimists? First, it is central to our species’s mentality that we look for explanation. Even though all explanation must bottom out at some point, we value explanation that puts things into more informative terms, terms that help us to connect things we don’t understand to things we understand better. This drive underwrites the proliferation of stories, myths, theories, models, parables, and diagrams designed to explain our world—it’s what we do. Likely, we’re incurable explanatory optimists in this respect. But the optimism I’m invoking has its more recent roots in the Scientific Revolution. It is here that the sort of systematic explanatory success relevant to thinking about the hard problem emerges.

Explanatory optimism stems from the Scientific Revolution, which greatly expanded our understanding of the natural world. In the centuries since the advent of modern science in the early 1600s, more and more seemingly disparate phenomena have been explanatorily incorporated into our scientific theories. Of course, great mysteries and challenges remain, but as a trend, this progression should spark some optimism about our ability to illuminate any particular natural phenomenon in scientific terms. Or at least we should be reasonably slow to pessimistically declare something inexplicable in these terms.⁷ Explanatory optimism is an attitude, a theoretical bet about the possibility of progress in our investigations of the natural world.⁸ History suggests that recalcitrant explanatory puzzles—those seemingly marking the limit of our explanatory reach—often end up, with enough time and effort, being properly located and illuminated in the framework of modern science.

But what is this framework and how does it lead to optimism about explanation in specifically structural, dynamical, and functional terms? Perhaps there is good reason to be optimistic about scientific explanation in general, but this attitude does not extend to the more narrowly defined idea of explanation in structural, dynamical, or functional terms. I contend that optimism extends even to the more narrow characterization. To begin, there is an overarching background recommendation we should take from looking

at the successful history of scientific explanation: we should embrace “methodological naturalism.” Methodological naturalism is the idea that, as Daniel Stoljar puts it, “the best methods that we have for finding out about the world are the methods, whatever they are, that one finds in the natural sciences” (2010, 11).⁹ The explanatory successes since the Scientific Revolution strongly recommend the doctrine, especially in this relatively thin form. Below, I will have more to say about what methodological naturalism is (and isn’t). For now, we can see it as the suggestion that where we can, we ought to employ the explanatory methods of modern science, barring decisive barriers to their application.

And what, broadly, are the methods of modern science? This is, of course, an enormous and controversial question, but to make my optimistic point, it will be sufficient to isolate several general and wide-ranging trends, all again having their origins in the explanatory shifts marking the emergence of modern science. A central innovation is the trend toward mechanistic explanation. Mechanism has developed considerably since its early modern form, but it has some basic core commitments we can extract.¹⁰ A first idea takes the natural world to be analogous to a machine (most notably a giant clock), one whose behavior can be explained in terms of the causal interaction of basic, inanimate parts. Mechanistic explanation in this sense involves the push and pull of component elements. The human body, nonhuman animals, celestial motion, and the weather were all seen as mechanical systems, amenable to decomposition and explanation in terms of the interaction of causally intertwined parts. The extension of this sort of thinking to an ever-widening range of phenomena is part of the success of modern science.

Another element of mechanistic explanation downplays the distinction between natural and artificial systems, holding that so long as the right causal processes are instantiated, we get an explanation of the system in question. Common to this thread is the idea that nature is composed of basic bits of matter in motion and that these together allow for the explanation of all phenomena, natural or artificial. This suggests an ontology of basic elements or atoms or “corpuscles” characterized by how they causally interact to produce observable phenomena. This facet of mechanical explanation is implicated in the rejection of various Aristotelian ideas (for example, substantive form and teleology), a rejection symptomatic of the Scientific Revolution. The specific details of various early scientific mechanistic theories are not important for my purposes. Rather, together such approaches support the theoretical claim that all we need to explain complex, high-level phenomena is an understanding of the arrangement of basic constituents, constituents characterized solely in terms of their quantifiable role in producing the appropriate motions. They also point to a long-term trend toward the development of quantifiable mathematical and geometric models in the sciences.

These two explanatory elements—a corpuscular approach and “mathematicalization”—emerged with the rise of mechanical explanation (broadly conceived) at the outset of the Scientific Revolution. Together, they account for many of the famous developments of modern science, leading to Newton’s theory and beyond. The trend toward mathematicalization is present in Galileo’s seminal work developing and defending the methods of the “new science.” In *The Assayer*, Galileo compares nature to a great book:

But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth.

(Galileo 1623/1957, 237–238)

This approach is picked up and extended by such figures as Descartes and Newton, and it represents a clear mark of modern scientific methodology, particularly as it played out in physics. The scope, predictive power, and fruitfulness of such mathematically rooted theories highly recommend the idea that we should employ such methods where we can to explain natural phenomena. There is controversy over how far this mathematical approach can be extended into domains other than physics, and there is a deep question about whether mathematical models are idealized fictions or whether they pick out the true nature of basic reality. But, for our purposes, we can see that explanation in mathematical terms, as in Newton’s or Einstein’s theories, or appropriate connection to such explanatory frameworks, is to be sought out where feasible. Further, the employment of “corpuscular” theorizing lends itself to mathematicalization. If complex higher-level phenomena are explained in terms of simpler lower-level elements, elements fully characterized in terms of a set of quantities like position, acceleration, number, and change of position, then there is a clear route to laying out a mathematical model of these interactions. Show how the quantifiable corpuscles causally interact over time and you show how the whole system works. Capturing natural phenomena in this sort of explanatory framework is a key achievement of modern science.

However, in the contemporary literature, the strands of “clockwork mechanism” and “mathematical mechanisms” (as we can term them) are often seen as pulling in different explanatory directions. To the point, scientific practice in the special sciences may not in fact proceed by pursuing mathematically specifiable law but may instead look for decomposable causal processes involving stable systems with interacting parts. But, for my purposes, both trends point to explanation in structural, dynamical, or functional terms. Clockwork mechanism is a paradigm of explanation in terms of the pushing

and pulling of interacting elements without appeal to any further powers of the composing parts beyond their role in producing the causal output of the mechanism in question. There is still the question of the nature of the fundamental bits composing the parts of the mechanism, but if a non-fundamental phenomenon can be explained in clockwork mechanistic terms, then there is only call to posit those properties at the fundamental level needed to account for causal processes.¹¹ And explanation in terms of mathematized physical law again leaves us features specified in terms of relational properties—those properties specifiable in terms of quantifiable connections and interactions (ultimately cashed out in mathematical terms). There is no appeal to anything beyond structural, dynamical, and functional properties in the relevant sense. Thus, both brands of mechanical explanation lead to optimism in favor of explanation in structural, dynamical, and functional terms.

We are left with good reason to be optimistic about explanation in structural, functional, and dynamical terms. Needless to say, there are a range of complexities I have not addressed in making this case—more precise characterizations of the kinds of properties and processes at issue, the nature of mechanism and mathematical modeling, the proper story about explanation, and so on. But I think we can step back from these details and retain the key point: the history of successful scientific explanation, building in pace over the last 100 years, has been achieved in “easy problem” terms, the very sorts of terms that explanatory pessimists believe cannot tell the full story of consciousness.

We are therefore faced with a choice-point in our theorizing. We have good reason to be explanatory optimists. But first-person access to consciousness gives us reason to doubt that the smooth run of easy explanation can continue.¹² First-person access seems to provide evidence that consciousness is (at least partially) constituted by properties going beyond structural, dynamical, or functional properties. We can call these “exceptional” properties, the properties underwriting the hard problem.

Exceptional properties: Properties underwriting the hard problem by being inexplicable in structural, dynamical, or functional terms

If consciousness possesses exceptional properties, optimism fails. There are two main options before us. We can accept that things are as first-person access presents them and embrace the pessimistic conclusion. We then proceed to craft our worldview in light of this explanatory limit. We posit new and exceptional features of reality, ones that will transcend the limits of past science. Perhaps in time a new sort of explanatory framework will develop, or maybe this is as far as our understanding can go when it comes to consciousness. Various forms of dualism, panpsychism, and nonexplanatory emergent physicalism fall into this category. The second option, on the other hand,

denies that things really are as first-person access presents them to be. Explanatory pessimism is rooted in the deliverances of first-person access. We seem to be in close, perhaps infallible, epistemic contact with exceptional properties. But if this contact misleads us about the true underlying nature of conscious experience, then explanatory optimism is still on the table. Further, if we have good reason to think we've been led astray by first-person access, reason provided by plausible empirical assumptions and evidence about the brain, then optimism is the proper default. This second path is the one I'll pursue in this book. It involves undercutting the purported evidence of first-person access with regard to the nature of conscious experience—the apparent evidence that consciousness involves exceptional properties. This, in turn, demands a debunking story, one showing that we have good reason to believe that things would appear in just this way even though everything is unexceptional. I will spend the rest of this chapter providing more detail about the sort of undercutting debunking argument I'll be employing. The following chapters will then fill out the debunking claim with theoretical and empirical detail, ultimately justifying an optimistic conclusion.

1.2 Undercutting and Debunking

When faced with what seems to be something apparently standing outside our usual picture of things, we can react in two ways. The first is to react metaphysically: we can posit changes to our current scheme, adding new elements. The second is to react epistemically: we can question whether we are right to think there really is an exception at all. Perhaps what seemed to be an exception is really just more of the same old. Most responses to the hard problem involve some kind of metaphysical reaction. We accept that we are indeed in the presence of something new and move accordingly. “Naturalistic dualists” posit non-physical exceptional properties connected to the physical by way of special psycho-physical laws. “Epiphenomenal dualists” hold that non-physical exceptional properties float outside the causal action, although they are real and needed to explain conscious experience. “Panpsychists” theorize that exceptional properties cannot be explained in more basic terms but still hold that they are in some sense physical, occupying the intrinsic, fundamental core of all physical matter. Or, perhaps as “neutral monists” hold, there is some more basic something, neither mental nor physical, lying at the most fundamental level of existence, and it accounts for the exceptional properties we find in conscious experience. Or we may allow that there are indeed exceptional properties in consciousness but argue (with “identity theorists”) that we have good reason to identify them with higher-level physical properties, perhaps properties of the brain. If an identity theorist accepts that there are exceptional properties, they count as embracing a metaphysical reaction on my view, even if they ultimately identify such