Is Consciousness Everywhere?
Essays on panpsychism

Edited by Philip Goff and Alex Moran
Contents

Title Page .................................................. i
Publisher Information ................................. ii
About the Authors...................................... iii

Is Consciousness Everywhere? ....................... 1
  1. Introduction ........................................... 1
  2. The Scientists ........................................ 3
  3. The Philosophers ..................................... 5
  4. The Theologians ..................................... 7
  5. Replies from Philip Goff .......................... 8

1: The Scientists ......................................... 9
  Consciousness and the Laws of Physics .......... 10
  Relations and Panpsychism ......................... 30
  Physics, Time, and Qualia ......................... 35
  The Real Problem(s) with Panpsychism .......... 55
  Reflections of a Natural Scientist on Panpsychism ... 72
  Autism and Panpsychism ............................. 85
  Dr Goff, Tear Down This Wall! ................... 105
  What is a Theory of Consciousness for? ....... 121

2. The Philosophers ..................................... 136
  Is Panpsychism at Odds with Science? .......... 137
  A Solution to the Combination Problem .......... 153
  Galileo's Real Error ................................ 168
  Qualities and the Galilean View ................. 175
  Grounding the Qualitative ......................... 195
  Panpsychism and the Limits of Physical Science .. 217
  Missing Entities .................................... 232
  Can a Post-Galilean Science of Consciousness... 253
  'Oh You Materialist!' ................................ 273

3. The Theologians ..................................... 298
  Why a Panpsychist Should Adopt Theism .......... 299
  Panpsychism and Spiritual Flourishing .......... 321

4. Replies from Philip Goff .......................... 346
  Putting Consciousness First ....................... 347

Also Available ........................................ 395
About the Authors

**Damian Aleksiev** holds a PhD in philosophy from the Central European University. He works primarily on the metaphysics of consciousness and questions of fundamental ontology. He is currently working on a project that explores the use of explanatory gaps as a guide to reality.

**Sean Carroll** is an American theoretical physicist who specializes in quantum mechanics, gravity, and cosmology. He is a Research Professor in the Walter Burke Institute for Theoretical Physics in the California Institute of Technology Department of Physics and an External Professor at the Santa Fe Institute.

**Marina Cortês** is a physicist who works at Berkeley Lab in California. She completed her PhD at the University of Sussex. Research interests include physics of the early universe, Bayesian statistics, dark energy reconstruction and phenomenology, and gravity of discrete space times.

**Jonathan Delafield-Butt** is Professor of Child Development and Director of the Laboratory for Innovation in Autism at the University of Strathclyde. His work examines the origins of conscious experience and the embodied and emotional foundations of psychological development. He took his PhD in Neurobiology at the University of Edinburgh, extended to Developmental Psychology at the Universities of Edinburgh and Copenhagen. He has trained pre-clinically in psychoanalytic psychotherapy.
Chris Fields is an independent scientist working at the intersection of quantum information theory, evo-devo biology, and cognitive neuroscience. His primary interest is in how systems at all scales identify objects in their environments, represent state changes over time, and develop communicative conventions that enable cooperative and/or competitive activity. Fields has published over 190 refereed papers in various areas of physics, cognitive science, and biology.

Keith Frankish is Honorary Reader in Philosophy at the University of Sheffield, Visiting Research Fellow at the Open University, and Adjunct Professor with the Brain and Mind Programme in Neurosciences at the University of Crete. Keith has wide research interests in philosophy of mind and cognitive science, and he has published numerous articles, books, and edited collections, including Illusionism as a Theory of Consciousness (Imprint Academic 2017).

Philip Goff is an Associate Professor of Philosophy at Durham University. He defends panpsychism as the best account of how consciousness fits into our overall theory of reality. Goff has authored an academic book, Consciousness and Fundamental Reality (OUP), and a book aimed at a general audience, Galileo's Error: Foundations for a New Science of Consciousness. He has published 45 articles as well as writing extensively for newspapers and magazines.


Christof Koch is a neuroscientist best known for his writings on the basis of consciousness, starting with the molecular biologist Francis Crick over 25 years ago. Trained as a physicist, Christof was for 27 years a professor of biology and engineering at the Caltech.
In 2015, he became the President of the Allen Institute for Brain Science in Seattle. He’s now the Chief Scientist of the MindScope Program at the Allen Institute. He’s also the Chief Scientist of the Tiny Blue Dot Foundation, with its focus on understanding consciousness.

**Joanna Leidenhag** is Lecturer of Theology and Liberal Arts at the University of Leeds. She earned her PhD from the University of Edinburgh in 2018. Her first monograph, *Minding Creation: Theological Panpsychism and the Doctrine of Creation*, was published in 2021 by T&T Clark. Her research includes theological evaluations of panpsychism, as well as theological engagement with autism and mental health.

**Michelle Liu** is a Leverhulme Early Career Fellow at the University of Hertfordshire where she also worked as a lecturer in philosophy (2019–21). She completed her DPhil at Oxford in 2019. She works in philosophy of mind, philosophy of language, and aesthetics, and is particularly interested in phenomenal experience, pain, polysemy, and the expressiveness of art.

**Alex Moran** is a Leverhulme Early Career Fellow at the University of Oxford. His PhD was completed at the University of Cambridge (Queens’ College). His primary research interests are in metaphysics and the philosophy of mind. Moran also has research interests in metaethics, early analytic philosophy, and early modern philosophy.

**Alyssa Ney** is Professor of Philosophy at the University of California, Davis. She is the author of *The World in the Wave Function* (OUP 2021), *Metaphysics: An Introduction* (Routledge 2014), and editor, with David Albert, of *The Wave Function: Essays in the Metaphysics of Quantum Mechanics* (OUP 2013). She works primarily on the interpretation of quantum mechanics and the unity of science.
Robert Prentner completed a PhD in physical chemistry and, later, in philosophy at ETH Zürich. He did his postdoc studies on the interface theory of perception at UC Irvine, and worked as Senior Research Fellow at the Center for the Future Mind at Florida Atlantic University. His research pertains to the philosophical foundations in the sciences of mind, in particular mathematical consciousness science. He is co-editor of the journal *Mind and Matter* and co-organizer of the ‘Models of Consciousness’ conference series. He is currently based at the Munich Center for Mathematical Philosophy at LMU Munich.

Sarah Lane Ritchie is Lecturer in Theology and Science at the University of Edinburgh. Her research is in science and religion, with a particular focus on the intersection of cognitive science of religion, psychology, theology, philosophy of mind, and metaphysics.

Luke Roelofs is a postdoc at NYU’s Centre for Mind, Brain, and Consciousness, where they work on issues in the metaphysics of consciousness and the philosophy of imagination. Their first book, *Combining Minds*, seeks to defend panpsychism against its much-discussed ‘combination problem’.

Carlo Rovelli is a theoretical physicist known for his work in quantum gravity. He founded the Quantum Gravity group of Aix-Marseille University. He is member of the Institute Universitaire de France, Honorary Professor of the Beijing Normal University, Honoris Causa Laureate of the Universidad de San Martin, Buenos Aires, and member of the Académie Internationale de Philosophie des Sciences. He is currently Adjunct Professor of Philosophy at the Western University and Distinguished Visiting Researcher at the Perimeter Institute in Canada.

Anil K. Seth is Professor of Cognitive and Computational Neuroscience at the University of Sussex, Co-Director of the Canadian Institute for Advanced Research (CIFAR) Program on Brain, Mind and Consciousness, an ERC Advanced Investigator,
Editor-in-Chief of the journal *Neuroscience of Consciousness*, and a Wellcome Trust Engagement Fellow. He has published more than 180 papers. His TED talk has been viewed more than twelve million times, and his new book, *Being You: A New Science of Consciousness*, is an instant *Sunday Times* Top 10 Bestseller and *Guardian* book of the week.

**Lee Smolin** is a theoretical physicist who has done most of his work on quantum gravity. He has co-founded a number of approaches including loop quantum gravity and relative locality. He has broad interests and has contributed new ideas to cosmology (cosmological natural selection), foundations of quantum mechanics (non-local hidden variables theories, energetic causal sets, etc.). He is the author of five semi-popular books that explore the philosophical ramifications of the big open questions in physics. Since 2001 he is a founding and senior faculty member of Perimeter Institute.

**Galen Strawson** holds the President’s Chair of Philosophy at the University of Texas at Austin. His books include *Mental Reality* (1994) and *Selves* (2009). His most recent books are *The Subject of Experience* (2017) and *Things That Bother Me* (2018).

**Clelia Verde** is the product development manager at an innovative insurance company based in Milan. Trained as an engineer (MSc: Risk Management, MIB School, Trieste, Italy), she has over the years consistently pursued her interests in advanced physics and the literary arts. She is a contributing science editor for an independent Italian general-interest news site (Gli Stati Generali) and an award-winning poet (Bookcity Milano 2019).

**Ralph Stefan Weir** is Lecturer in Philosophy at the University of Lincoln and Associate Member of the Faculty of Theology and Religion at the University of Oxford. He received his PhD in philosophy from the University of Cambridge in 2020. His recent publications include ‘Does Idealism Solve the Problem of Consciousness?’ in *The Routledge Handbook of Idealism and Immaterialism*, ‘Christian Physicalism and the Biblical Argument
1. Introduction

The mind–body problem, broadly speaking, is the challenge of understanding how the conscious mind relates to the physical world. On the one hand, it seems that there is nothing more familiar to us, nothing we know better, than the phenomenon of consciousness. On the other hand, there is much about consciousness that has defied our understanding. In particular, there seems to be a special difficulty when it comes to reconciling the subjective and qualitative aspects of consciousness with our objective and quantitative scientific picture of the physical world. One major task for science and philosophy, therefore, is to find a way to bridge this gap, thereby explicating the place of consciousness in nature.

Historically, two main positions have been defended in answer to the mind–body problem. One is dualism, the view that


[2] Durham University, UK. philip.a.goff@durham.ac.uk

[3] University of Oxford, UK. alexander.moran@philosophy.ox.ac.uk
consciousness lies outside the physical domain, and therefore has a wholly non-physical nature. The other is materialism (or physicalism), the doctrine that, since reality is wholly physical in nature, consciousness must ultimately be thought of as a part of the material world. The present volume, however, focuses on an alternative theory; namely, panpsychism. Like the dualist, the panpsychist claims that consciousness is irreducible and fundamental, and hence cannot be understood in other, more basic, terms. Like the materialist, however, the panpsychist also thinks that consciousness does not lie outside of the rest of nature, but is rather firmly located within the material world. Indeed, the panpsychist makes the radical claim that consciousness is ubiquitous in nature, in so far as it is a property instantiated, not just by humans and some animals, but even by the most fundamental constituents of physical reality. The panpsychist thus takes nature to be permeated by consciousness, rather than viewing consciousness as a derivative phenomenon that only emerges at higher levels.

One central thought motivating the panpsychist position is an observation about our understanding of the physical world itself. The key idea is that physics only describes the relational or dispositional properties of matter, not its intrinsic nature: what matter does rather than what it is. This then raises a question about what the intrinsic nature of matter actually is. According to the panpsychist, the intrinsic nature of basic matter is constituted by rudimentary kinds of consciousness. This then allows us – or so it is hoped – to explicate the more familiar phenomenon of human consciousness in terms of the more basic kind of consciousness instantiated by fundamental physical things.

Proponents of panpsychism argue that the theory enables us to make significant headway with the traditional mind–body problem, while avoiding the well-known problems that dualists and materialists face. Its detractors, meanwhile, urge that the theory faces substantial problems of its own. The papers in this special issue explore these and related issues, from the perspectives of science, philosophy, and theology. Some papers focus on further
motivating and developing the panpsychist position. Others explore various challenges that the panpsychist faces. Collectively, they shed new and important light not only on panpsychism, but on the fundamental question of the place of consciousness in nature more generally.

As stalking horse, many of the papers focus on Philip Goff’s book *Galileo’s Error: Foundations for a New Science of Consciousness*, which offers an important and accessible defence of the panpsychist view. The special issue also includes a response piece from Goff to these various articles.

2: The Scientists

Our first three papers are by theoretical physicists. Carlo Rovelli has previously defended a relational interpretation of quantum mechanics, according to which quantum mechanics concerns not how physical entities are in themselves, but how they are *in relation to one another*. In his paper, Rovelli suggests that this relational conception of physics is itself a – very mild – form of panpsychism; and moreover that it may help address the intuitions underlying the ‘hard problem’ of consciousness, as the differences between the mental and the physical are now less stark than they might previously have appeared.

The next two papers share a common focus. Panpsychists believe that consciousness is a fundamental feature of reality. Both of the next two papers (the first by Sean Carroll, the second by Marina Cortês, Lee Smolin, and Clelia Verde) agree that this conviction requires rewriting current physics, as the understanding of the basic physics in brains given to us by the ‘Core Theory’ (the standard model of particle physics combined with the weak-field limit of general relativity) does not make reference to consciousness. From this point of agreement, they go in different directions. Carroll infers that consciousness is not a fundamental feature of reality, but rather an emergent property of certain complex systems. Cortês, Smolin and Verde, in contrast, present a new interpretation of
fundamental physics in which qualia – as well as the passage of time – play a fundamental role.

The next two papers are by neuroscientists, one pro panpsychism and one opposed. Anil Seth believes that, rather than focusing on the ‘hard problem’ of where consciousness comes from in the first place, it is more profitable to focus on the ‘real problem’ of explaining, predicting, and controlling the various properties of consciousness in terms of physical processes in the body and brain. Moreover, he rejects panpsychism as an untestable and therefore unfruitful hypothesis. Whilst Christof Koch is much more sympathetic to panpsychism, like Seth, he emphasizes the importance of experimental science and testable predictions. Koch outlines and defends the integrated information theory of consciousness, a theory which entails that consciousness is ubiquitous in the physical world and hence can be seen as a form of panpsychism. Koch also expresses disagreement with Philip Goff’s exegesis of Galileo in *Galileo’s Error*.

We often forget that there are philosophical assumptions lying at the bedrock of any scientific worldview. In his work as an experimental psychologist, Jonathan Delafield-Butt has found that dropping the philosophical assumptions of materialism, and adopting in their place panpsychist assumptions, affords deeper insights into the nature and character of autism. In his paper, Delafield-Butt lays out the case for this, with reference to the panpsychist framework of Alfred North Whitehead.

In our next paper, Robert Prentner defends the *interface theory*, which he has developed in collaboration with Donald Hoffman and others. The interface theory has much in common with panpsychism. Both take fundamental reality to be made up of consciousness. However, whereas panpsychism holds that the physical world is also fundamental (because the physical world is made up of consciousness), the interface theory – or ‘idealism’ as we philosophers have called it for a couple of hundred years – holds that there is a more fundamental (mental) reality *underlying* the physical world. Prentner pits panpsychism against the interface
theory, arguing that the latter offers a more robust, less dualistic theory of consciousness.

Chris Fields has developed a detailed form of panpsychism: minimal physicalism. In his paper in this special issue, however, Fields raises some challenges to Goff’s conception of the science of consciousness. Fields is sceptical that science can or should be in the business of accounting for the specific character of conscious states, e.g. the redness of a red experience, as opposed merely to accounting for why those conscious experiences exist at all (it is interesting that this approach seems to be the precise opposite of the ‘real problem’ approach Seth defends in his paper). Fields also expresses his suspicions that the real motivation for the ‘consciousness war’ between dualists, materialists, and panpsychists may be a yearning, on the part of some opponents of materialism, for human exceptionalism.

3. The Philosophers

The special issue includes nine papers by academic philosophers (excluding the response piece from Goff). Two of these papers defend and elaborate the panpsychist position. Luke Roelofs explores the question as to whether, and to what extent, panpsychism classifies as a scientific worldview. In addition, the paper offers a nuanced discussion of what exactly it means for a philosophical thesis such as panpsychism to count as scientific in the relevant way. In the following paper, Annika Harris explores a fundamental problem that panpsychists face, which has gained much attention in the recent literature, known as the ‘combination problem’. Harris diagnoses the problem as resulting from the conviction that, for any given conscious experience, there exists a ‘subject’ or a ‘self’ that has the experience. Once we realize that subjects don’t really exist in the first place, she argues, the problem goes away.

Five further papers focus on raising challenges for panpsychism; a couple also promote an alternative, physicalist view. Damian Aleksiev argues that, while they focus on explaining consciousness,
panpsychists face problems when it comes to accounting for the physical world. In particular, the argument is that if we conceive of the intrinsic nature of matter in terms of consciousness then several facts about physical reality become difficult to explain. Alyssa Ney, meanwhile, argues that panpsychism is not, contrary to what its defenders maintain, sufficiently well-motivated by its underlying claims concerning physics and the nature of matter. She also argues that physicalists are in just as good a position as panpsychists when it comes to accounting for free will, objective value, and meaning.

The remaining three essays from this grouping challenge panpsychism from a different angle. Again, panpsychists are primarily concerned with explaining consciousness. However, as these authors point out, there is a related explanatory task, concerning, not consciousness, but the sensory qualities of external things (such as the redness of a rose, or the distinctive smell of coffee). Both Keith Frankish and Michelle Liu argue that panpsychists repeat the ‘Galilean mistake’ of supposing that the sensory qualities are instantiated ‘in the mind’ rather than by external objects. Frankish then argues that, instead, we should deny that anything instantiates such qualities, which, he claims, then makes it that much easier to embrace a reductive materialist position. Michelle Liu, by contrast, defends sensory quality realism, i.e. the view that external things really possess the full range of sensory qualities they seem to, and criticizes the panpsychist for not respecting this position. In a similar vein, Moran presupposes precisely the kind of sensory quality realism that Liu defends, and then maintains that, on this assumption, there is much about the physical world that panpsychists are not in a position to explain. He also claims that, while certain nearby positions to panpsychism may be able to meet the challenge he articulates, there is a certain kind of non-reductive physicalism that fares just as well.

The two final papers by philosophers are concerned with connections between panpsychism and the more familiar materialist and dualist positions. While sympathetic to Goff’s proposal for a ‘post-Galilean’ science of consciousness, in which
consciousness is taken to be a fundamental feature of reality, Ralph Weir argues that post-Galileans end up being committed to a radical form of dualism known as substance dualism. Pushing in a contrary direction, Galen Strawson urges that panpsychism in fact classifies as a form of materialism, and indeed arguably the most defensible form. Strawson also provides a helpful, detailed discussion of what the doctrine of materialism actually amounts to.

4. The Theologians

Panpsychism is often assumed to be a spiritual doctrine. However, many contemporary proponents of panpsychism are resolutely secular. They may not believe in a transcendent reality but they do believe that people have experiences – they feel pain, they see colour – and this mundane and everyday reality needs to be accounted for in our overall theory of reality.

Nonetheless, some have argued that there are important connections between panpsychism and spiritual convictions of some form or another. Joanna Leidenhag, in her paper, argues that the motivations that lie behind the arguments put forth in support of panpsychism, if applied consistently, lead to belief in God. The panpsychist demands an intelligible account of how consciousness emerges. But is it consistent to demand an explanation of consciousness without also demanding an explanation of the existence of the universe itself? Leidenhag thinks not, and suggests this line of reasoning can be satisfactorily concluded only by a commitment to theism.

In the final essay of the volume, Sara Lane Ritchie explores the connections between panpsychism, psychedelic experiences, and spiritual flourishing. It is common for the person undergoing a psychedelic experience to feel a connection to ultimate reality, however that is conceived (Ritchie examines a number of options, including panentheism, the view that ‘the entire natural world exists within God, but also that God is, in some sense, more
than the natural world’). Presumably a materialist must reject these experiences as delusional. However, Ritchie argues that the worldview of a panpsychist may be consistent with the veridicality, i.e. the non-delusional nature, of these kinds of psychedelic experiences.

5. Replies from Philip Goff

In the final article, Philip Goff responds to the essays of the volume. He also explores some ideas on what a ‘post-Galilean’ science of consciousness – one which takes consciousness to be a fundamental feature of reality – might look like.
1: The Scientists
Consciousness and the Laws of Physics
Sean Carroll

Abstract: We have a much better understanding of the dynamics and ontology of physics than we do of consciousness. I consider ways in which intrinsically mental aspects of fundamental ontology might induce modifications of the known laws of physics, or whether they could be relevant to accounting for consciousness if no such modifications exist. I suggest that our current knowledge of physics should make us sceptical of hypothetical modifications of the known rules, and that without such modifications it’s hard to imagine how intrinsically mental aspects could play a useful explanatory role.

1. Introduction
We don’t fully understand consciousness. That’s hardly surprising. The human brain, which is at least somewhat involved in consciousness, contains roughly 100 billion neurons and 700 trillion synaptic connections. It is arguably the most complex structure in the known universe. Even as neuroscience makes impressive advances in understanding the brain, it seems prudent to anticipate that we have a number of conceptual and technical

[4] California Institute of Technology and Santa Fe Institute, USA. seancarroll@gmail.com

breakthroughs yet to come that could bear in important ways on the question of consciousness.

We do, on the other hand, understand the basic laws of physics governing the stuff of which brains are made. They take the form of an effective quantum field theory describing a particular collection of matter particles interacting via force fields. There is certainly much of physics remaining to be discovered, but in the specific regime covering the particles and forces that make up human beings and their environments, we have good reason to think that all of the ingredients and their dynamics are understood to extremely high precision (Carroll, 2021a). Modern physics, in other words, provides evidence for what philosophers call ‘causal closure of the physical’: physical events have purely physical causes (Loewer, 1995; Papineau, 1995), at least in the regime relevant to human life. Without dramatically upending our understanding of quantum field theory, there is no room for any new influences that could bear on the problem of consciousness.

Given this situation, it might seem surprising to a disinterested observer to learn that anyone would argue that the best route toward understanding consciousness involves augmenting or altering the ontology suggested by fundamental physics. To start with the least-well-understood aspects of reality and draw sweeping conclusions about the best-understood aspects is arguably the tail wagging the dog. When we can’t remember where we put our car keys, we don’t typically respond by going out and buying a new car.

Nevertheless, a prominent strain in the philosophy of consciousness proposes to do just that (Chalmers, 1996; Goff, 2017; 2019). The justification for such a radical move is that there will be something qualitatively missing in any account of consciousness based purely on physical ontology as we currently understand it. This perspective arises from a conviction that physics can explain behaviour, but not the first-person experiences characteristic of human consciousness; that physics may account for the dynamics of the stuff in the universe, but it doesn’t illuminate the intrinsic nature of that stuff.
In this paper I support the idea that physics is in such good shape that the most promising strategy for trying to understand consciousness is as a (weakly) emergent phenomenon that leaves physical ontology untouched, rather than trying to extend or elaborate that ontology with specifically mental aspects (cf. Moran, this issue; for a contrary view see Cortês, Smolin and Verde, this issue). After reviewing the ‘Core Theory’ and our reasons for being confident in its accuracy, I will discuss what it would mean to modify it, either directly in the dynamics or by adding additional ontological features. It is always possible that contemporary physics is inadequate and in need of modification, but a close examination highlights the difficulty of doing so in a rigorous and convincing way.

2. The Physics Underlying Everyday Life

Science often employs multiple vocabularies or theories for describing the same physical situation, often at different degrees of focus or coarse-graining. These are often called ‘levels’, although strictly speaking they need not be arranged hierarchically. Within any level, we can specify the domain of circumstances in which a particular theory is applicable. The claim here is that there is one level of description – that of effective quantum field theory – and a well-defined regime – interaction energies below certain thresholds, broad enough to include every situation encountered in ordinary human life – where we have very good reasons to believe we know precisely what is going on.

The fundamental ontology of any quantum theory is specified by a ‘quantum state’ or ‘wave function’, expressed mathematically as a vector in an abstract Hilbert space (Carroll, 2021b). In a quantum field theory, that state can be thought of as being constructed from possible configurations of fields that take on values at each point in space-time.

Fortunately, the details of this formalism are not necessary for our present purposes. Once we quantize the fields,
appropriate configurations – essentially, low-lying energy states – can be interpreted as collections of interacting particles. These circumstances are more than broad enough to encompass human beings and their environments. Thus, we can think of people and the objects around them as configurations of certain particles. In particular, human beings are made of atoms; those atoms are made of protons, neutrons, and electrons; the protons and neutrons are made of quarks and gluons. These particles interact through gravitation, electromagnetism, and the nuclear forces, and get mass from a background Higgs field.

The dynamics of these particles and forces are governed by an effective quantum field theory known as the ‘Core Theory’, consisting of both the standard model of particle physics and the weak-field limit of general relativity (Wilczek, 2015). This theory is not the ultimate theory of everything, nor is it intended to be. The world might not be described by a quantum field theory at the deepest level; that description might emerge from a more fundamental set of degrees of freedom and dynamical laws. And the Core Theory is certainly not supposed to cover every circumstance – dark matter and the Big Bang, to name some obvious examples, are not included. But we have excellent reasons to believe that the entirety of the ‘everyday life regime’ supervenes on the ontology and dynamics of this theory (Carroll, 2021a). If there is a more fundamental level, its properties are irrelevant to the autonomous dynamics of the Core Theory. And if there are additional particles and forces, they interact too weakly with the known fields to exert any influence on human behaviour; otherwise they would have already been detected in experiments.

Our confidence in this picture derives from the fact that quantum field theories are the practically unique way to satisfy the general principles of quantum mechanics and relativity; from symmetries ensuring that any unobserved fields must be too weakly-interacting with ordinary matter to be relevant for everyday-life dynamics; and the property of effective field theories that the dynamics themselves are fully determined in terms of a very small number of parameters.
We can’t know for certain that the Core Theory suffices to correctly describe the behaviour of the particles and fields making up human beings, no matter how good our arguments become, but any proposed modification of this theory should be held to a very high standard indeed. Just as with any hypothetical new physical model, it should be quantitative and precise, detailing exactly how the explicit dynamics of the Core Theory are meant to be modified, and how such modifications are consistent (or not) with features such as unitarity, locality, symmetries, and conservation laws, not to mention experiments.

3. Domains of Applicability

In the context of the relationship between consciousness and the laws of physics, it is worth being a bit more explicit about how we specify the ‘domain of applicability’ of a theory (Carroll, 2016). The general idea is that there is a set of physical situations in which the predictions of the theory are meant to be accurate, with no claims being made for situations outside that set.

The empirical foundation of the Core Theory has been established through a line of experimental and observational results stretching back to Faraday, Rutherford, and many others. But the most precise constraints come from modern-day particle colliders, which typically measure the results of scattering individual particles off of each other. One might sensibly wonder whether results from such a paradigmatically reductionist setting can be straightforwardly extrapolated to something as complex as a human brain, which contains roughly $10^{27}$ particles. Perhaps brains are just not within the domain of applicability of the Core Theory.

If we accept the basic framework of effective quantum field theory, this concern is unfounded; everything that happens inside biological organisms here on Earth is unambiguously within the purview of the Core Theory. Its domain of applicability is bounded by two criteria. The first is that gravity must be weak, so that we can treat the gravitational field as an ordinary quantum field,
sidestepping subtleties of horizons and Hawking radiation. ‘Weak’ is a relative term, and in this case means ‘the gravitational potential is much smaller than one’. In practice, this means ‘we are nowhere near a black hole’. This criterion is easily met by everything we know of in the Solar System, human brains included.

The other criterion comes from effective field theory. The modifier ‘effective’ indicates that the domain of applicability of the theory is specified in terms of energies – in particular, the amount of energy transferred between particles when they interact. An effective field theory is meant to be accurate when energy transfers remain lower than some explicit cut-off. In the case of the Core Theory, experiments have established its accuracy at energy transfers of up to $10^{11}$ electron volts. Electrochemical reactions inside biological organisms, meanwhile, happen at less than $10^2$ eV. Shrinking the domain of applicability of the Core Theory while remaining within the framework of effective quantum field theory requires a mistake in our current understanding by a factor of over a billion, which seems implausible.

The effective field theory paradigm also features very specific properties of the field dynamics: they are local (interacting only with other fields at the same space-time point), and governed by a simple and inflexible set of equations. So below, when I refer to ‘within the effective field theory paradigm’, this is what is meant: a theory of quantum fields, evolving under the appropriate simple dynamical equations, applicable in circumstances where gravity is weak and interactions feature energy transfers below the cut-off.

Within its domain of applicability, the Core Theory is what we might label \textit{causally comprehensive}. If we give a complete specification of the quantum state of the Core Theory fields within that regime, there is a specific equation that unambiguously predicts how it will evolve over time. This equation is sufficient to describe everything human beings generally do, unless they jump into a black hole or stick their hand inside the beam of a high-energy particle accelerator. There are no ambiguities or loose ends. The fact that brains are big, complex things is irrelevant. The Core
Theory makes specific predictions for how any particular brain will behave; our choice is to either accept that prediction, or modify the theory in some way. There is no third alternative (Aristotle, 2002).

4. Ontology and Dynamics

Despite the extraordinary empirical success of the Core Theory, and the fact that human beings and their brains are made out of particles interacting within its domain of applicability, there is a lingering worry that no physicalist picture is up to the task of accounting for consciousness, even as some higher-level weakly-emergent phenomenon. There are various ways of expressing this concern: conscious experiences are inherently first-personal and subjective; merely physical objects cannot feel what it is to be like something; describing the behaviour and functions of objects does not explain their intrinsic nature; and others. See Goff (2019) for an overview.

One common reaction to these concerns is to contemplate modifications of the underlying ontology suggested by modern physics: to suggest that a quantum state built upon interacting fields obeying strict equations of motion is incapable in principle of accounting for consciousness, and that we instead need to add specifically mental aspects to our description of reality. We may contemplate ontological modifications as dramatic as substance dualism, in which an immaterial mind is distinct from the physical body but interacts with it, or idealism, in which the physical world is a kind of projection of a fundamentally mental reality. I will focus on more subtle approaches, in which mental aspects or properties are related to, but augment, the basic physical reality. Approaches under this umbrella include property dualism, which posits distinct mental properties in addition to physical properties (Chalmers, 2003); Russellian monism, which posits both physical and mental aspects belonging to a single underlying set of properties (Russell, 1927; Chalmers, 1996; Strawson, 2006; Goff, 2017); and other forms of panpsychism, epiphenomenalism, and
related approaches (Papineau, 2020). For convenience I will refer to any new ontological features as ‘mental aspects’, which is meant to include potentially autonomous properties as well as intrinsic qualities that might supervene on the physical situation.

Any such approach must specify whether, and how, it modifies the dynamics of the theory as well as the ontology. In our conventional understanding, consciousness exerts an important influence on behaviour: I can have a conscious experience and talk about it. (Admittedly, highly trained philosophers are reported to be able to imagine removing consciousness from a being without affecting its behaviour in any way.) Observed human behaviour can be traced to electrical and chemical signals in our brains and nervous systems. Explicitly mental aspects of ontology could affect this behaviour by, for example, influencing the rates of chemical reactions, or the strength of electromagnetic forces, or the probability of certain quantum outcomes.

In what follows we will examine different kinds of relationship that a theory of consciousness might have to the physical dynamics of the Core Theory, as well as the possibility that there is no relationship at all. Our goal is not to comprehensively catalogue the possibilities, but just to highlight some of the challenges faced by any approach that aspires to explain consciousness by adding mental aspects to the fundamental ontology of the world.

5. Consciousness and Quantum Mechanics

Quantum field theory is a subset of quantum mechanics. Like any quantum theory (and in contrast with classical theories), the dynamics of the Core Theory come in two parts. There is a law of evolution that describes how an undisturbed quantum state evolves deterministically over time, referred to as the unitary dynamics. The other part takes the form of a probabilistic algorithm expressing how the wave function responds to being measured. Operationally, a measured wave function ‘collapses’ onto a state with a definite value of the quantity being measured, so we label this the collapse
dynamics. Collapse introduces a stochastic element, with the probability of different outcomes being related to the original wave function by the Born rule. There are therefore two broad strategies one could contemplate for modifying the dynamics of the Core Theory: altering its unitary dynamics, or its collapse dynamics.

The questions of what precisely constitutes a quantum measurement, what happens when one is performed, and what is the correct ontology describing quantum systems have remained controversial. We don’t need to distinguish between these competing theories for our present purposes; see overviews by Norsen (2017) and Maudlin (2019), and cf. Rovelli (this issue).

According to textbook quantum mechanics, when one measures a quantum observable such as position or momentum or spin, only certain specified outcomes can be obtained. To each possible outcome, the quantum state assigns a complex number, the amplitude. The Born rule states that the probability of obtaining that outcome is the modulus-squared of the corresponding amplitude. Importantly, there is no hidden structure within this rule; once we know the amplitudes, experimental outcomes are truly randomly chosen from the appropriate probability distribution.

The Born rule has thus far passed experimental tests (Jin et al., 2017), but the fact that both consciousness and quantum measurement remain mysterious makes it tempting to imagine that there is a connection. What we are interested in here is not the prospect that consciousness causes wave-function collapse (Wigner, 1961; Stapp, 2001; Chalmers and McQueen, 2014), but that somehow wave functions collapse in just the right way to account for consciousness. Penrose and Hameroff have developed an approach in which wave functions collapse when certain physical criteria are met, which they argue can explain aspects of human cognition (Penrose, 1989; 2014; Penrose and Hameroff, 2011). However, although this programme is often described as an approach to ‘consciousness’, it does not attempt to answer the qualitative questions of first-person experience any differently than any other purely physical account. Similarly, quantum
entanglement may play a role in cognition (Fisher, 2015), but this is a matter of information processing, without any special connection to qualitative experience.

If one were interested in allowing mental aspects to affect the probability of quantum measurement outcomes, presumably that could be done. The Born rule states that the probability of obtaining an outcome $a$ is given by $p(a) = |\psi_a|^2$, where $\psi_a$ is the component of the wave function corresponding to that outcome. We could imagine a new rule

$$p(a) = f(\psi_a M_a)$$

(1)

where $M_a$ represents some novel mental aspect of the situation. This modified Born rule might affect the rate of certain chemical reactions inside a human brain, thereby allowing mental aspects of consciousness to influence our physical behaviour, without showing up in experiments performed with non-conscious equipment.

Of course, such a rule for wave-function collapse represents a wild modification of conventional physics, not merely a loophole within it. A respectable theory along these lines would include a specification of what the mental aspects $M_a$ are, an understanding of their independent dynamics, and an explicit form of the new rule (1). All of these are possible to contemplate, but they remind us of the high standards to which any modified laws of fundamental physics should be held.

Furthermore, if one were convinced that purely physical ontologies are incapable in principle of accounting for the qualitative features of consciousness, the process of wave-function collapse does not offer any unique opportunities. Regardless of when and how wave functions collapse, at the end of the day they are still wave functions. If we think in terms of novel mental properties affecting chemical reaction rates, there would be no relevant difference between modifying the collapse dynamics and the unitary dynamics.
6. Consciousness and Quantum Field Theory

We turn next to the unitary dynamics. As discussed above, if we stay entirely within the effective field theory paradigm, both for the Core Theory fields and potentially new dynamical elements, there is no room for modifying the dynamics in ways that would be relevant for human behaviour while remaining compatible with experimental constraints. Any new fields that would be relevant for what goes on in the human brain would have been discovered long ago.

We can nevertheless imagine that new mental aspects influence the quantum fields of the Core Theory, without themselves obeying the rules of quantum field theory. To see how that might work, it is useful to look at one part of the Core Theory: quantum electrodynamics, the theory of charged particles (including electrons, protons, and even atomic nuclei) with electromagnetic fields. To the extent that gravity, nuclear reactions, and radioactive decays can be ignored, this is enough to include all of the physics relevant for human biology. The unitary dynamics can be summarized in a one-line equation:

$$A = \int_{k<\Lambda} [DA][D\psi] \exp \left\{ i \int d^4x \left[ \sum_n \bar{\psi}_n (i\gamma^\mu \partial_\mu + q_n \gamma^\mu A_\mu - m_n) \psi_n - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \right] \right\}$$  \hspace{1cm} (2)

We don’t need to dive into this equation in detail, but a few points are worth highlighting. The expression tells us how a quantum state (describing, for example, a human brain) consisting of charged particles $\psi_n$ and electromagnetic fields $F_{\mu\nu}$ evolves from a given initial state to a final one. It is entirely deterministic and causally comprehensive; indeterminism only comes from the non-unitary collapse dynamics. The expression in square brackets is the Lagrangian, which encodes the properties of different kinds of particles. The notation $\int d^4x$ indicates that the Lagrangian is integrated over space-time. This reflects the locality of the unitary dynamics: fields only
interact with other fields (and with themselves) at the same point in space-time. The notation \( k < \Lambda \) indicates that this is meant to be an effective theory, applicable below the cut-off energy \( \Lambda \). See (Carroll, 2016) for further elaboration. Any given approach to consciousness will either modify this equation, or it won’t.

The properties specified by the Lagrangian include the masses of the particles \( m_n \), as well as the parameters characterizing their interactions, such as electric charges \( q_n \). The strength and rate of electrochemical processes, including those in human brains and bodies, are calculable in terms of these parameters.

An obvious way that mental aspects could modify physical dynamics is for them to affect the values of these parameters, which would in turn affect the rate of processes in the brain. Given some physical/mental situation \( S \), we could imagine context-dependent changes in the values of the physical constants that govern Core Theory dynamics, of the form

\[
m_n \rightarrow m_n (S), \quad q_n \rightarrow q_n (S).
\]

If \( S \) included mental aspects of our ontology, this would be a mechanism by which those aspects could affect human behaviour, such as our testimony concerning our introspective experiences.

As in the case of the Born rule, we are welcome to contemplate mentally-induced modifications of particle-physics parameters, but a number of questions present themselves. What precisely is meant by the situation \( S \), and what kind of dynamics does it have? Naïvely, changes in the masses or charges of particles would lead directly to violations of conservation of energy and momentum. Are these compensated by transfers of energy between conventional matter and a ‘mental sector’? These questions are potentially answerable, but they highlight the challenges faced by any proposed theory of this form.
7. Mental Degrees of Freedom

Panpsychists sometimes analogize consciousness to electric charge, as a property that inheres in appropriate fundamental particles. There are at least two severe limitations to this analogy. First, electric charge is a paradigmatic example of a property with dynamical consequences: placed in an electric field, particles with opposite charges move in opposite directions. In the case where we imagine that the properties associated with consciousness have no dynamical consequences, it is not clear what the analogy is supposed to illuminate. Second, charge is conserved. An elementary particle has a single, unchanging value of charge throughout its existence, whereas it is generally supposed that conscious states can take on different values.

We should therefore distinguish between two alternatives: that hypothetical new mental aspects of our ontology supervene on the physical situation, or that there are independent ‘mental degrees of freedom’ that are not determined by the physical situation. In panpsychist terms, these correspond to the possibility that any particular electron has a definite value of consciousness, versus the idea that any given electron might have multiple conscious states.

If the new mental aspects of our ontology supervene on the physical aspects, as far as physical behaviour is concerned this is indistinguishable from not introducing mental states at all. Consider some causal chain $P_i \rightarrow M_i \rightarrow B_i$, where $P_i$ is some physical state, $M_i$ is some mental state, $B_i$ is some behaviour, and ‘$\rightarrow$’ stands for ‘inevitably leads to’. If we think of mental aspects as primary, but nevertheless supervening on the physical situation, we could write $M_i(P_i) \rightarrow B_i$. Either case is functionally equivalent to the shorter chain $P_i \rightarrow B_i$. For all intents and purposes this is equivalent to positing that mental aspects have no effect on physical behaviour.

Turn instead to the alternative where the same physical configuration might be associated with different mental degrees of freedom. These aspects would be roughly analogous to the spin of an electron, which is some combination of ‘spin-up’ and ‘spin-
down’ for any given particle, but can be specified independently of the particle’s position.

The idea of new independent mental degrees of freedom runs into immediate trouble. In conventional field theory, the existence of new degrees of freedom quantitatively affects processes that rely on quantum fluctuations, in which each property value represents a separate contribution that should be added together (e.g. we ‘sum over spins’ in a scattering calculation). But we know empirically how many degrees of freedom actual electrons have – two spin states for the electron, and another two for its antiparticle, the positron. If electrons could also be found in both ‘happy’ and ‘sad’ states, it would have an unmistakable impact on their scattering rates, in flagrant contradiction with experiment.

Our allowed alternatives are to posit that all electrons have the same conscious state, in which case there is effectively no dynamical impact, or that mental degrees of freedom are somehow not like physical ones. In the latter case, we are left with the question of what mental degrees of freedom are like, if they are not like physical ones. We can avoid conflict with what we know, but only at the expense of pushing our ideas further away from clarity and tangibility.

8. Strong Emergence

One approach to the relationship between consciousness and physics is to appeal to strong emergence – the idea that legitimately new behaviours arise in collective phenomena that cannot be derived in terms of the individual behaviours of constituent parts of the system. Strong emergence is sometimes invoked as a way to allow for specifically mental causal powers (O’Connor and Wong, 2005). It is worth spending a moment on the relationship between strong emergence and the underlying framework of effective quantum field theory – namely, they are entirely incompatible.

As discussed above, the Core Theory provides a comprehensive specification of the quantum-field dynamics within its domain
of applicability, which includes any processes between known particles with energy transfers less than a hundred billion electron volts. The field equations are precisely local: the unitary dynamics of each field at any one space-time point are influenced only on the values and derivatives of the other fields at the same point, and not directly by what is happening elsewhere. Electrons and other particles obey the same equations whether they are inside a rock or inside a human brain.

The strong emergentist must therefore deviate from the paradigm of effective field theory entirely, while maintaining the empirical successes of the Core Theory. The most direct way to do this would be to postulate a new restriction on the domain of applicability that is not given in terms of energy transfers in particle interactions, but on some explicitly macroscopic criterion. For example, one could hypothesize that quantum field theory breaks down when the number of particle excitations in a region surpasses a certain number, or when the configuration of such particles reaches a certain quantifiable degree of complexity or information processing capacity. The effective masses and couplings of elementary particles might, for example, be modified as in equation (3), where the situation $S$ could involve a quantitative measure of consciousness from an approach such as integrated information theory (Tononi et al., 2016).

One is, of course, free to contemplate whatever extravagant deviations from contemporary physics one likes. Particle-physics experiments typically examine the interactions of just a few particles at a time, so new physical laws that only kick in for complex agglomerations of particles are not necessarily ruled out by data we currently have. It’s worth noting, however, how profound a departure such laws would represent. The most fundamental principle of quantum field theory is locality: fields at any one point in space-time are only influenced by the values and derivatives of other fields at that same point, not the behaviour of fields at other points. Modifying the dynamical equations in ways that
were sensitive to the complexity of a configuration of surrounding particles would represent a dramatic overthrow of this principle.

Moreover, based on purely physical grounds rather than consciousness-based motivations, our expectation that the laws of quantum field theory might break down in biological organisms would be very low indeed. To we macroscopic people, the $10^{27}$ particles in a human brain seems like a lot, certainly far greater than the number physicists typically collide in high-energy accelerators. But the density of those particles is very low by particle-physics standards. To be conservative, we might take as a standard length scale the Compton wavelength of the electron, $\lambda_e = 2 \times 10^{-10}$ cm. The volume of a human brain is 1,260 cubic centimetres, or about $10^{32}$ cubic Compton wavelengths. The number of particles is therefore less than $10^{-5}$ per standard volume. From the point of view of particle physics, a brain is not a densely packed system; indeed, it’s practically empty space. There is no physical rationale for expecting the dynamics of the Core Theory to break down in such an environment, regardless of how complex the overall situation is. For any particular electron or nucleus, almost all of the rest of the brain is so far away as to be essentially irrelevant.

This is not to say that the concept of strong emergence might not be relevant in other contexts, where the ‘micro’ theory is something other than elementary particles. If a complex system consists of a collection of smaller systems that are themselves complex, a purely local theory might not suffice, and the best description of the overall dynamics could conceivably involve microphysical dynamics that depend on macrophysical contexts in interesting ways (Flack, 2017). In the phenomenon of ‘quorum sensing’, for example, gene expression in bacteria is affected by their overall population density (Miller and Bassler, 2001). In such cases, the subsystems themselves are extended objects with non-trivial internal dynamics, so the criterion of locality is significantly less severe. Quantum field theory is a very different situation, where subsystems (elementary particles) have no internal structure. In that case, the locality of interactions is exact; what matters to the
dynamics of a field at each point is only the other fields at that same point, not anything elsewhere. Any new context-dependent behaviour departing from the predictions of equation (2) would be a violation of our expectations from effective quantum field theory, not a supplement to them.

9. Conclusions

The temptation to augment the ontology of the world with specifically mental aspects stems from a conviction that describing the mere behaviour or function of matter cannot be sufficient to account for consciousness or innate nature. As characterized by Levine (1983), there seems to be an ‘explanatory gap’ between physical states and conscious experiences. Physicalism posits that a conscious experience is an emergent phenomenon that arises in higher-level models of the same underlying processes described by physics. To a panpsychist, as Goff (2019) says about the brute identity theory, this ‘is very unsatisfying’. Arguably it is this ‘satisfaction gap’, more than any explanatory or ontological gap, that prompts the introduction of intrinsically mental concepts and categories into fundamental ontology.

Any discussion of mental aspects of ontology must specify one of two alternatives: changing the known laws of physics, or positing that these aspects exert no causal influence over physical behaviour. We cannot rule out the first option either through pure thought or by appeal to existing experimental data, but we can ask that any modification of the Core Theory be held to the same standards of rigour and specificity that physics itself is held to. The point of expressions like (1) and (3) is not that mentally-induced modifications of physical parameters are impossible, but that a promising theory of consciousness should be specific about how they are to be implemented.

We don’t know everything there is to know about the laws of physics, and there is always the possibility of a surprise. But the solidity of our confidence in the Core Theory within its domain
of applicability stands in stark contrast with our fuzzy grasp of the nature of consciousness. The most promising route to understanding consciousness is likely to involve further neuroscientific insights and a more refined philosophical understanding of weak emergence, rather than rethinking the fundamental nature of reality.

Acknowledgments

It is a pleasure to thank Philip Goff for stimulating this work and for enlightening discussions. I’d also like to thank Jenann Ismael, Barry Loewer, Alex Moran, David Papineau, Alex Rosenberg, and Eric Schwitzgebel for helpful comments and pointers to references. This work is supported in part by the Foundational Questions Institute.

References


Relations and Panpsychism
Carlo Rovelli

Abstract: Twentieth-century physics has revealed a pervasive relational aspect of the physical world. This fact is relevant in view of some of the motivations for panpsychism. In fact, it may be seen as a vindication of the panpsychist idea of a monist continuity where some aspects of consciousness’s perspectivalism are universal. On the other hand, this same fact may undermine some of the motivations for more marked forms of panpsychism.

If some aspects of mind are universal, which ones are so? I point out in this note that twentieth-century physics has already vindicated a – very mild – form of panpsychism. This is because of the profoundly relational aspect of physics, manifest in general relativity, but especially in quantum mechanics. Twentieth-century physics is not about how individual entities are by themselves. It is about how entities manifest themselves to one another. It is about relations.

This is particularly evident for quantum theory, especially (but not uniquely) if one reads it in terms of its relational interpretation (Rovelli, 1996; Laudisa and Rovelli, 2021). Niels Bohr expressed this ‘contextual’ aspect of quantum theory by saying that a physical system can only be described taking into account the systems it is

[6] Centre de Physique Théorique de Luminy, Aix-Marseille University, France. rovelli.carlo@gmail.com

interacting with. The current account of the world in fundamental physics is therefore always the account of how a system affects another system.

This implies that the most effective way of thinking about the world is not in terms of entities with properties, but rather in terms of systems that have properties in relation to other systems. In turn, this implies that any physical description of a system is necessarily perspectival: relative to another system.

In the textbook Copenhagen formulation of quantum theory, the physical system with respect to which properties take value is variously and a bit obscurely interpreted as ‘the apparatus’, ‘the observer’, ‘the macroscopic world’, and similar. In relational quantum mechanics, the properties of a system $S$ are defined with respect to any other physical system $O$ with which $S$ interacts, and are relative to $O$ (Laudisa and Rovelli, 2021). Similarly, in the many-worlds interpretation of quantum mechanics, variables have values only in relation to (components in a branch of) other systems when the two have got entangled (Saunders et al., 2010).

A direct consequence is that our physics today is not a physics of the world seen from the outside. It is a physics of the world always seen from the perspective of a physical system.

This is perhaps not panpsychism, because there is nothing specifically psychic or mental in the relational properties of a system with respect to another system. But there is definitely something in common with panpsychism, because the world is not described from the outside: it is always described relative to a physical system (Dorato, 2016). So, physical reality is, in our current physics, perspectival reality (Rovelli, 2021).

On the other hand, this very relationalism may suffice to resolve the very problems that motivated panpsychism in the first place: do we really need elementary physics to include more aspects in common with the mental world than this? Which ones? All phenomena of which I am aware that are more related to psyche or mind are connected to a brain, a neural system, sensory organs,