

# Putting Consciousness First

## Replies to Critics

I am honoured and humbled that these 21 incredible scientists and thinkers have responded to my work.<sup>1</sup> These essays have both challenged and stimulated me. There is much disagreement, as there should be in these matters about which there is little consensus. But there are also many of us who are journeying to the same location, albeit via different paths.

This also gives me an opportunity to sketch in more detail what a post-Galilean science of consciousness, one in which consciousness is taken to be a fundamental feature of reality, might look like. It's time non-reductionists about consciousness stopped spending all their time justifying their existence and got on with building an interdisciplinary research programme to rival the dominant materialist paradigm. In *Galileo's Error*, I argue that the problem of consciousness is rooted in the philosophical foundations of science. It follows that if we want to solve the problem, we need to rethink what science is. This will involve the joined-up efforts of physicists, philosophers, neuroscientists, and many others. In bringing together non-reductionists about consciousness from so many different fields, this volume takes an important step towards that goal. It's time to build the post-Galilean paradigm.

### Part I: Replies to Scientists

#### Reply to Carlo Rovelli

Carlo Rovelli's proposal is ingenious. However, I'm not persuaded it provides an adequate solution to the hard problem of consciousness. There are two aspects of consciousness that give rise to a hard problem: quality and subjectivity. Quality consists in the fact that experiences involve *qualities*: the smell of coffee, the taste of mint, the deep red you experience as you watch the setting sun. Subjectivity consists in the fact that these experiences are *for* someone: there is something that it's like *for me* to experience that deep red.

These two aspects of consciousness give rise to two 'hard problems.' Either would, in my view, be sufficient to refute materialism. But the hard problem of quality is more pronounced, which is why I tend to focus on it. Physical science tells us a purely quantitative story of causal structures. In that kind of vocabulary, you can't simply can't articulate the qualities of our experience. That's an *expressive* limitation of the language of physics. But I think it entails an *explanatory* limitation. Because if I wanted to reductively explain the redness of a red experience in purely physical terms, my theory would have to articulate the redness and then explicate it terms of more basic physical structures. If physical theory can't

---

<sup>1</sup> Unfortunately it was not possible to write a reply to Delafield-Butt's fascinating essay, due to last minute submission.

even articulate the redness, then it can't reductively explain it. (Of course, we can capture in quantitative terms the location of redness in the similarity space of colour along the dimensions of hue, lightness and saturation, but that kind of purely structural information doesn't fully convey the redness of a red experience).

The hard problem of subjectivity is a very similar challenge, but more subtle. You can't articulate in the language of causal structures the hypothesis that a certain physical system has *experience*. Many materialists have tried to get around this by articulating in the language of causal structures something that *sounds a bit like* experience. We can talk, for example, about how a certain physical system is 'sensitive' to certain features of the environment, say, the red colour of objects, in the sense that it's causally set up to track red objects or respond in some way to the fact that things are red. That sounds a bit like saying the system is experiencing red. But of course it's not the same thing at all. The fact that a system causally-mechanically responds to a certain feature of reality doesn't entail that it has any experience of that feature. Another popular option has been to talk about the capacity of system to monitor its internal states, which can obviously be captured in causal-structural terms. That sounds a bit like the system is 'aware' of its internal states. Well, it depends what you mean by 'awareness'. If you just mean 'causally responds to', then, yes, the system is aware. But if you mean that the system 'experiences' its internal states, then this is no way implied by its causal responsiveness to its inner states.

Rovelli's position is a highly novel form of this old strategy of describing something that sounds a bit like experience, and then suggesting that this solves – or at least helps with – the hard problem. As Rovelli says, his favoured interpretation of quantum mechanics is essentially 'perspectival.' Having a 'perspective' seems to imply having experience, or something on the way towards it. But this is just a reflection of ambiguity in the word 'perspective'. We sometimes say that something 'has a perspective' to mean that there is something that it's like to be it, that it *experiences*. But when in relational quantum physics we say that a system 'has a perspective', we mean something quite different: that the theory cannot be applied to systems in isolation but only in relation to each other. Accepting that physical systems 'have perspectives' in the relational quantum physics sense is totally consistent with physical systems not 'having perspectives' in the experiential sense.

I don't think Rovelli would deny this. After all, he's not saying his view is literally a form of panpsychism. But if we haven't managed to articulate, and thereby explain, subjectivity in the terms of physical science, then we haven't addressed the hard problem. And even if relational quantum mechanics could close the 'subjectivity gap' – the gap between the processes of physical science and the having of experience – this would still leave the 'qualitativity gap' – the gap between the quantitative features of physical science and the qualities of experience – as wide as ever.

I'm not sure Rovelli would disagree with this either, as much of his article is devoted to rejecting the idea that there is a hard problem in the first place, arguing that 'what we can "conceive" depends on the conceptual structure we have, and this keeps changing and includes a big deal of presuppositions, sometimes wrong.' I am totally on board with this when it comes to anything other than consciousness. The unique thing about consciousness science is that we have a fundamental explanandum that does not come from public

observation and experiments, but from the immediate awareness each of us has of the qualities of our experience (as well as the fact that we are experiencing those qualities). Introspection is fallible in all sorts of ways. However, the basic concepts we use to articulate the fact that we experience, and that our experience involves qualities, are not subject to scientific revision in the way the concepts of time, space and solidity are. Why not? Because we know that we have experience, and that our experience involves qualities, with a greater certainty than we know any empirical fact.

Of course, one could reject all this, and follow Daniel Dennett in holding that there is no first-person explanandum for a science of consciousness, that the only task for any kind of science is to explain the facts of public observation and experiment. But in that case, there is no hard problem, and hence no need to employ relational mechanics to address the hard problem. At the end of the day, I don't see stable middle ground between Dennett and Chalmers.

### Reply to Sean Carroll

Sean Carroll worries that any view according to which consciousness is a fundamental feature of reality will end up requiring modifications to a well-confirmed scientific theory, namely the Core Theory. It is more plausible, therefore, to adopt a view on which consciousness is a weakly emergent feature of reality. Panpsychism comes in two forms: weak and strong emergentist. I dispute that either version has the implication Carroll accuses it of. Let us take each version in turn.

#### **Weak emergentist panpsychism**

The Core Theory describes the causal dynamics of fields and particles, as determined by their physical properties, such as mass, spin and charge. According to Russellian panpsychism, those physical properties are forms of consciousness. Hence, according to panpsychism, the Core Theory describes the causal dynamics of fields and particles *as determined by the forms of consciousness they instantiate*. Rather than physics not leaving room for consciousness to have an impact, the entire story of physics is the story of what consciousness does. Of course, when you're doing physics, you don't know that's what you're studying. But that's just because physics is only concerned about causal dynamics and abstracts away from the nature of the things underlying those dynamical structures. Doing physics is like playing chess when you don't know what the pieces are made of.

Carroll says panpsychists 'analogize' consciousness to electric charge. That's not quite right. According to panpsychism, charge *is* a form of consciousness. Carroll objects that '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.' The fact that electric charge has dynamical consequences is totally consistent with its being a form of consciousness; it just follows that that form of consciousness has dynamical consequences, it does stuff. Carroll also worries that the charge of an elementary particle is unchanging throughout its entire existence. But the experience of a human being is constantly changing due to the constantly changing processes of the brain; there is nothing incoherent in the idea of a very simple mind with unchanging experience.

Carroll then raises two possibilities: either there is a one-to-one correspondence between physical properties and elementary forms of consciousness or there isn't. The Russellian panpsychist will certainly opt for the former, given that physical properties just are forms of consciousness. But Carroll now worries that the panpsychist is proposing the following causal chain:

Physical property *causes* mental property *causes* behavioural impact

Or

Physical property (upon which mental property supervenes) *causes* behavioural impact.

In either case, Carroll claims, this is 'functionally equivalent to' and '[f]or all intents and purposes...equivalent to..':

Physical property causes behavioural impact.

The true panpsychist view is captured by neither of these. There is no causal relation between the physical property and the mental property, as they are one and the same thing. And if anything, the supervenience relation goes in the other direction: a certain form of consciousness is correctly classed as 'charge' in virtue of its dynamical consequences. Carroll is right that these descriptions are 'functionally equivalent,' if that just means that from the perspective of a science that is only interested in behaviour, they come out the same. But to assume that such a science of behaviour captures the full story of what is going on is to beg the question against panpsychism. According to panpsychism, physics is so useful because it just focuses on behaviour, abstracting away from the real nature of the properties being studied.

### **Strong emergentist panpsychism**

Strong emergentist panpsychism, as the name suggests, combines panpsychism with the kind of strong emergence Carroll discusses in his contribution. As strong emergentists, they believe that certain complex systems, such as conscious brains, have novel causal capacities that could not be predicted from knowledge of their basic components. Imagine a superintelligence, of the kind imagined by Simone Pierre Laplace, who has total knowledge of the particles and fields covered by the Core Theory at time  $t$ , and tries to work out the state of my brain at  $t+1$ , solely on the basis of the Core Theory. If strong emergence is true, that superintelligence will make some false predictions about the locations of the particles in my brain at  $t+1$ , as it is relying entirely on the Core Theory and is ignorant regarding the contribution of the emergent causal capacities of my brain.

Crucially, however, this does not entail that strong emergentists are modifying the Core Theory. They can instead take the Core Theory to be complete on its own terms, i.e. as a complete theory of the inherent causal capacities of particles and fields. However, if there are strongly emergent wholes, these wholes themselves have irreducible capacities, capacities which *complement* the basic causal capacities of particles and fields. When these strongly emergent wholes come into being, they *co-determine* the evolution of the universe,

in conjunction with the basic causal capacities of particles and fields (where the latter are perfectly captured by the Core Theory).

Carroll is right that the strong emergentist is obliged to do some serious theoretical work. But this theoretical work need not be conceived of as modifying the Core Theory, but rather as explaining how the causal capacities of strongly emergent wholes interact with the causal capacities of particles/fields to co-determine what will happen. Understanding strong emergence in this way gives us a response to Carroll's novel argument that '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.' Maybe so, but we should think of strong emergence not as quantum field theory breaking down but as a new neuro-biological theory kicking in. And the place to look for when emergent neurobiological principles kick is not physics but neurobiology.

It seems to me an open empirical question whether or not strong emergentism is true. Some philosophers (Papineau 2001) have put forth an inductive argument against strong emergence: if there was strong emergence in the brain, we probably would have found it by now, therefore, it probably doesn't exist. In my early work, I assumed the soundness of this kind of argument, and used it as part of a defence of weak emergentist panpsychism. However, the more I talk to neuroscientists, the more I'm inclined to think the first premise of this inductive argument is false, as we simply don't know enough about the workings of the brain. We know a great deal about the basic chemistry of the brain: neurotransmitters, action potentials, calcium channels, etc. And we know a fair bit about the large-scale functions of the brain. However, we know very little about how these large-scale functions are realised at the cellular level (Cobb 2020), i.e. about how the brain *works*. Because we have little clue about how the functions of the brain are realised, we have little clue about whether they are entirely realised by the basic electro-chemical processes we understand so well, or whether at some point new causal principles kick in which, in conjunction with the electro-chemistry, realise the brain's functions.

People get very excited about brain scans, but in fact they are very low resolution. Each pixel of an fMRI image corresponds to 5.5 million neurons, between 2.2 and  $5.5 \times 10^{10}$  synapses, 22km of dendrites, and 220km of axons (Logothetis 2008). We are only 70% of the way through putting together a complete connectome of a maggot's brain, with its 10,000 neurons (Cobb 2020: 257). The idea that we know enough about the workings of the human brain with its 86 billion neurons to know whether or not its workings involve strong emergence is not credible.

As we uncover more about the workings of the brain, if there does turn out to be strong emergence, this would provide a crucial way of making progress on identifying the neural correlates of consciousness (NCC). Michel Mattias (2019) has argued that for 150 years the science of consciousness has been having the same debates about where to locate consciousness, without significantly narrowing down the options. In my view, the root of these difficulties is that consciousness is not a publicly observable phenomenon. We are totally reliant on the external signs of consciousness, such as report. But the 'detection procedures' which underly inference from external markers of consciousness to consciousness itself are controversial, and arguably impossible to justify empirically. For

example, there is much debate over so-called ‘overflow’: the thesis that there are conscious experiences we are unable to attend to. If the overflow thesis is true, then it does not follow from the fact that somebody reports not having a given experience that they did not in fact have that experience. One manifestation of these difficulties is the perennial dispute (Odegaard et al 2017, Boly et al 2017) over whether consciousness is located at the back or the front of the brain.

But if there is strong emergence, then it will show up empirically. There will be functions of the brain that are not realised solely in the underlying electro-chemistry. To over-simplify: suppose it turned out that these irreducible functions were found in the front but not the back of the brain. This would provide powerful evidence that the seat of consciousness is located in the front of the brain.

It is notable that there are scientists exploring non-reductionist models. The neuroscientist Kevin Mitchell’s (2019) working assumption is that there are strongly emergent dynamics in the brain, and he seeks ways of modelling these emergent dynamics. And Martin Picard’s Mitochondrial PsychoBiology Laboratory model the behaviour of mitochondria in the brain as social interactions rather than attempting to reduce them to underlying chemistry. These are the kind of research programmes non-reductionists about consciousness can build on in shaping the post-Galilean alternative to the materialist paradigm.

#### Reply to Marina Cortes, Lee Smolin and Clelia Verde

Much of my reply to Carroll above doubles as a reply to Cortes, Smolin and Verde. Like Carroll, Marina Cortes, Lee Smolin and Clelia Verde interpret Russellian panpsychism as holding that consciousness is ‘not involved in dynamical laws,’ and hence ‘it can’t matter one whit whether they are there or not.’ But, as I responded to Carroll, Russellian panpsychists do not hold that the physical properties do all the work whilst micro-consciousness sits along for the ride. The Russellian view is that micro-consciousness is *all that exists* at the fundamental level, and hence micro-consciousness is doing the work if anything is. In referring to ‘mass’, ‘spin’ and ‘charge’, physics is – unbeknownst to itself – referring to kinds of micro-consciousness. Furthermore, as I also press in my response to Carroll, even if biological consciousness does make a fundamental causal impact on the universe, that doesn’t entail modifying fundamental physics.

Having said all that, this is a fascinating proposal. Cortes, Smolin and Verde are doing precisely what Carroll suggests is required of a strong realist about consciousness: modifying physics in order to make space for consciousness to play a fundamental role in the causal evolution of reality. I would love to see the empirical implications spelt out in more detail, especially those pertaining to the neural correlates of consciousness. I discussed in my reply to Carroll how strong emergence may allow us to make progress on, say, the debates as to whether consciousness is in the back or the front of the brain. It’s an intriguing prospect that Cortes, Smolin and Verde’s proposal may be another way, e.g. if there turn out to be more events without precedent in the front of the brain. This is exactly the kind of interdisciplinary research programme we non-reductionists need to be formulating in building the post-Galilean paradigm. It’s also exciting that the post-Galilean approach may

lead to new insights in the ongoing attempts to understand quantum mechanics, or to reconcile quantum mechanics with general relativity (cf. my reply to Aleksiev below).

The other fascinating aspect of Cortes, Smolin and Verde's essay is the discussion of time. They worry about reducing temporal flow and causation to static, eternal truths. I'm a bit more agnostic on this issue. The reality of 'real time' is not an undeniable datum in the way the reality of consciousness is. I tentatively sketched an argument against four-dimensionalism (the view that all moments of time exist equally and hence time doesn't really pass) in my academic book *Consciousness and Fundamental Reality*. In the universe as the four-dimensionalist conceives of it, there is nothing which (a) persists through time, and (b) has the kind of conscious experience we pre-theoretically associate with humans and other animals. According to four-dimensionalism, each person is associated with a four-dimensional 'spacetime worm' stretching out over their entire life history. But if those spacetime worms are conscious, they have a very weird kind of consciousness consisting of a unified experience of a whole life. The momentary stages of the spacetime worms are the things that have 'ordinary' conscious experience, but these things are forever located at one temporal location and do not persist through time. This argument pressures the four-dimensionalist to deny a lot of common-sense, but I suspect many of them would be happy to do so (maybe not Carroll, who is keen to preserve the Manifest Image).

### Reply to Anil Seth

I'm so glad Anil Seth was able to contribute at the last minute. He gives a rigorous and compelling defence of a popular way of defending materialism, and this gives me a chance to clarify my opposition to materialism in response.

Consciousness is not a normal scientific phenomenon. In general, the aim of science is to account for the data of public observation and experiment. But in the case of consciousness, the phenomenon we are trying to account for is not publicly observable. You can't look into someone's head and see their feelings and experiences. We know that consciousness exists not from observation or experiment but from our immediate awareness of our own feelings and experiences. Of course, science is used to dealing with unobservables, but in all other cases, we postulate unobservables in order to explain what can be observed. In the unique case of consciousness, *the thing we are trying to explain* is not publicly observable.<sup>2</sup>

Moreover, the qualities we know about via immediate awareness of our experience cannot be described in the purely quantitative language of physical science. You can't convey to a colour-blind neuroscientist what it's like to see red. Do I just mean you can't get a colour-blind neuroscientist to have a red experience by reading neuroscience? No, it's more than that. There is *information* that we get from attending to our experience, information that cannot be conveyed in the language of physical science. That descriptive limitation entails an explanatory limitation, as I explain in my reply to Rovelli above.

---

<sup>2</sup> Mattias Michel (in our discussion on the *Mind Chat* podcast) suggested to be that this is no different to dinosaurs, who can't be publicly observed because they no longer exist. However, our grounds for believing dinosaurs existed consists entirely of publicly observable data. The crucial point in the case of consciousness is that *there are data that aren't publicly observable data*.

I therefore totally reject Seth's characterisation of the hard problem as simply the challenge of explaining how consciousness exists in the first place, as though everyone agrees that we have good materialist explanations of the qualitative character of specific experiences. No: in order to give a materialist account of the qualitative character of red experiences, you'd have to be able to convey that qualitative character in the language of physical science, and that simply can't be done (see my reply to Rovelli). If we assume a functionalist notion of representation, then predictive processing can help explain the representational content of experiences. But unless we can capture in the quantitative language of cognitive functions and predictive processing the qualities of our experience, then a reductive explanation of those qualities is impossible.

Am I suggesting that we can't deal with consciousness experimentally? Not at all. Although we can't publicly observe consciousness, we can ask someone what they're feeling and experiencing, or observe external markers of their consciousness. If we do this while scanning their brain, we can establish correlations between the neural correlates which we can observe and the private experiences we cannot. This is important data, but it's not a complete theory of consciousness. What we ultimately want is to explain *why* brain activity goes along with conscious experience. Because consciousness is not publicly observable, this is not a question one can answer with an experiment (all experiments can do is establish more correlations). At this point we must turn to philosophy, examining the various proposals philosophers have offered to account for the fact that brain activity is correlated with experience: materialism, dualism, panpsychism, and idealism are the standard options. (Or rather, we turn to what is currently philosophy but I hope it will one day be established science: philosophy is what exists when the rules of the game aren't agreed upon.)

Seth of course disputes this, holding that explanatory bridges can be built from publicly observable mechanisms to private experiences. But when you look into the details, it's the old 'bait-and-switch' trick Chalmers (1995) exposed twenty-five years ago. Seth starts off talking about the qualitative character of the experience, and then subtly moves to a purely functionalist notion of content (e.g. predictive processing), giving the impression they're the same thing.<sup>3</sup> Low and behold, a purely functionalist notion of content can be accounted for in functionalist terms!

The whole allure of the materialist line is rooted in the idea 'But look how well science has done!'. The analogy to life is supposed to support this. But the data relevant to life are all publicly accessible. Explaining privately accessible qualities is a totally different explanatory enterprise to anything else science deals with; that's why Galileo set it outside of the domain of science, allowing scientists to focus on what can be observed and quantified. We're now going through a phase of history in which physical science has gone so well that many are inclined to think it's the road to all truth, including a complete theory of consciousness. The irony is that physical science has gone so well precisely because Galileo put consciousness outside of its scope of enquiry. The fact that physical science has been able to reductively explain so much since it set aside the private qualities of experience gives

---

<sup>3</sup> To be clear, I'm not saying anything nefarious is going on here. These are difficult questions, and I have no doubt that Seth is as intent on getting the truth as I am.

us no grounds for thinking it will be able to reductively explain the qualities of experience themselves.

Seth thinks I'm setting the bar too high. Why think scientific explanations should be intuitive? It's nothing to do with what is or isn't 'intuitive'. The point is just that we need an explanation of the phenomenon. Either the qualities of experience are fundamental, or we can deduce them from more fundamental facts. You can't deduce the qualities of experience from the physical facts; therefore, we can't explain the qualities of consciousness in terms of the physical facts; therefore, materialism is false. Seth hasn't given us any grounds for doubting either of the premises of this argument.

Seth says that panpsychism isn't testable. But nor is materialism. Both panpsychism and materialism are philosophical theories that go beyond the correlations established by the experimental science of consciousness. The advantage of going for panpsychism is that you swap the hard problem for the combination problem, and the latter looks more tractable. We already have worked out theories (Mørch 2018, Goff 2021) that are both empirically adequate and eliminate explanatory gaps. Still, Seth exhorts us to 'wait and see' whether a method designed to explain quantifiable, publicly observable data will somehow explain privately accessible qualities, even though it's never had any success in that explanatory project and we have other explanatory paradigms that are already yielding insights.

Towards the start of his article, Seth says that panpsychism is unpopular among many academic researchers on consciousness, citing a survey at the ASSC. This is potentially misleading, as it implicitly identifies 'academic researchers of consciousness' with 'academic researchers working on the *experimental science* of consciousness.' Are philosophers not academic researchers? What Seth and I are arguing about here is the nature of the hard problem, and whether it is philosophically coherent to propose a materialist answer. This is a philosophical question and among academic philosophers working on that question, panpsychism is a well-respected view.<sup>4</sup> It's true that among philosophers and scientists specifically working on the *experimental science* of consciousness, materialism dominates. There's a job of work to be done communicating to the broader community that, among those working on these more theoretical questions in philosophy of consciousness, materialism no longer dominates in the way it once did.

Finally, like most of the scientists in this volume, Seth assumes that panpsychism requires rewriting physics. I've explained in my responses to Carroll and Cortes, Smolin and Verde why this misunderstands the view.

It's not either/or. We need the experimental science of consciousness. But, because consciousness is not publicly observable, we also need philosophical theorising. The latter without the former is lame but the former without the latter is blind. I'm sure this debate will go on, and I'm looking forward to talking at length with Anil when Keith Frankish and I host him on the *Mind Chat* podcast next month.

### Reply to Christof Koch

---

<sup>4</sup> We'll have more precise data on this when the new PhilPapers survey results are published this October.

I have learnt a great deal from my interactions with Christof Koch, and I take us to be comrades-in-arms in the same struggle. However, there are also important disagreements between us.

Koch questions the view of Galileo I present in *Galileo's Error*. This understanding of Galileo is not something I came up with, but one I have learnt from historians of science. Filip Buyse, for example, says the following:

The corporeal domain was limited by Galileo to the domain of primary properties, which could be described in mathematical terms alone. Man, however, could not be described in mathematical terms – or, at least, not completely. For Galileo, the being of man also included the domains of color, odor, doubt, joy, etc. Man was thus understood to distinguish himself from other beings precisely by his capacity to have these sorts of affections. Man, one could say, is a bundle of secondary qualities. (Buyse 2015: 34).

Given that I am assuming a pretty standard view in the history of science, I don't think it's correct to say I'm 'retrofitting [my] ideas of the mind onto Galileo.' However, Koch's article did press me to think harder about whether this familiar historical interpretation is correct.

Koch seems to imply that the Lockean distinction between primary and secondary properties has nothing to do with Galileo. However, Galileo makes a big deal of distinguishing the properties of 'material or corporeal substance' – size, shape, location, motion, number – from properties such as 'white or red, bitter or sweet, noisy or silent, and of sweet or foul odour,' which reside not in corporeal substance but in the 'sensible body' (Galilei 1623/2008: 185). So we certainly get something that looks very much like the Lockean distinction.

What did Galileo mean by 'sensible body' ('corpo sensitivo')? Koch suggests we should understand this in opposition to the Aristotelian notion of the sensitive soul (l'amina sensitive'). However, Galileo also uses the term 'l'amina sensitive' to describe the sensitive body, and the fact that Galileo says the qualities of consciousness are *not* in the corporeal body but *are* in the sensitive body would seem to suggest a kind of Aristotelian dualism. I therefore find plausible Buyse's interpretation that, 'by the use of the terms "sensible body", Galileo suggests that, besides the body, there is an Aristotelian, immaterial sensitive soul, which is linked with the body and which functions as a principle of distinction between human, animal, vegetative, and unliving bodies' (Buyse 2015: 33). Koch provides no counterevidence to this interpretation of Galileo as an Aristotelian dualist.

Combining this with Galileo's famous declaration that mathematics should be the language of natural philosophy, whose subject matter is corporeal substance, we can surmise that Galileo thought:

1. The qualities of consciousness do not reside in matter but only in the ensouled body.
2. Mathematics is the appropriate language for describing the properties of corporeal substance, the proper subject matter of natural philosophy.

If Galileo did think we could capture the qualities of consciousness could be captured in mathematics, why did he deny that they resided in corporeal matter?

In any case, my main aim in appealing to Galileo is to resist the common argument 'Physical science has done so well, of course it's going to explain consciousness.' Regardless of what Galileo thought, those of us who do think the qualities of consciousness resist mathematical analysis can argue that physical science has gone so well since Galileo showed physical scientists how to ignore these qualities. The fact that physical science has gone so well when it ignores the qualities of consciousness doesn't give us grounds for thinking it will do well when it comes to dealing with the qualities themselves.

Despite our common interest in panpsychism, Koch worries that philosophy ultimately gets us nowhere and the only way we're going to make progress on consciousness is by getting on with the experimental science. The subtitle of *Galileo's Error* is 'Foundations for a New Science of Consciousness.' I am not trying to do science, in this book, but rather to lay the foundations for a systematic methodology for investigating consciousness. I have explained in my reply to Seth above why I think dealing with consciousness essentially involves philosophical theorising as well as experiments and won't repeat those points here.

On the face of it, Koch seems to be proposing a *materialist* theory of consciousness: he suggests that conscious experience is identical with maximal  $\phi$ , which is a purely quantitative, physically realised property. In *Galileo's Error*, I reject such materialist accounts, because I don't think you can account for the *qualities* of consciousness in the purely *quantitative* language of mathematical physical science (see my responses to Rovelli and Seth above for further detail). Koch responds that IIT provides a counterexample to this claim:

Haun and Tononi...published a detailed, mathematical account how the phenomenology of two-dimensional space, say an empty canvas, can be fully accounted for in terms of intrinsic causal powers of the associated physical substrate, here a very simple, grid-like neural network.

I have no doubt that we can in principle map out the quantitative structure of visual experience in mathematical language. This is important work. But such a mathematical description cannot fully capture the qualities that fill out that structure. If it could, we could use the mathematical description to explain to a colourblind neuroscientist what it's like to see colour, which I think Koch would agree is absurd. As I argue above in my reply to Rovelli, this descriptive limitation of purely quantitative language entails an explanatory limitation. If a purely quantitative theory can't even convey the qualities of experience, then it certainly can't reductively account for them.

There's a certain irony in Koch saying we need to stop philosophising and get on with the science. If Daniel Dennett was saying this, I could understand. But Koch defends integrated information theory (IIT) and IIT is a highly philosophical theory. It employs contentious philosophical notions, such as 'intrinsic existence' and 'intrinsic causal power'; and it justifies itself not only through empirical investigation but on the basis of five axioms, supposedly

known on the basis of introspection, which are then translated into five postulates the theory claims are necessary and sufficient for consciousness. I'm not objecting to any of this; I think that we need to bring science and philosophy together to deal with consciousness. But I think Koch needs to accept that the theory he's defending is knee deep in philosophical assumptions and arguments. What's the difference between my claim that we can know through introspection that consciousness has qualities that can't be captured in purely quantitative terms, and Koch's claim that we can know through introspection that consciousness exists 'for itself'?

It might sound like Koch and I have strongly opposed views. In fact, this is the rivalry of small differences (likewise, I find I have bigger fights with friends on the same wing of politics than I do with friends on the other wing). Whilst I disagree with much of Tononi's philosophical framing of IIT,<sup>5</sup> I'm increasingly thinking the empirical core of IIT may be the only extant philosophically coherent theory of consciousness. This is because it is the only theory of consciousness able to provide *non-vague* physical correlates of consciousness, 'non-vague' in the technical sense of not admitting of borderline cases. I will finish by briefly outlining this thought.

There are fuzzy borderline cases between being 'tall' and being 'not tall'; one of the things I believe we know about consciousness in virtue of our direct awareness of it is that there cannot be fuzzy borderline cases between consciousness and non-consciousness. Something is either experiencing or it's not. IIT can account for such sharp boundaries: a system becomes consciousness at the precise moment it comes to embody more integrated information than its parts. Contrast with global workspace theory (GWT), according to which a system is conscious just in case it 'broadcasts' information to a wide variety of its cognitive systems, if it embodies 'fame in the brain,' to use Daniel Dennett's memorable phrase. How famous does the information need to be? Whatever level is deemed sufficient, there will inevitably be some fuzzy borderline cases between being famous enough and not famous enough, and hence some fuzzy borderline cases whereby there's no fact of the matter as to whether the system is conscious. Therefore, IIT passes, whereas GWT fails, a crucial test of a theory of consciousness.

The post-Galilean research programme needs to develop and institute these kinds of introspection-based tests of theories of consciousness. They may be crucial in making progress where purely third-person experimental methods break down due to the fact that consciousness is not publicly observable. The post-Galilean research programme will start to be noticed when it starts getting results.

#### Reply to Robert Prentner (and Donald Hoffman, et al)

I have a lot of time for the interface theory of consciousness – or 'idealism' as we philosophers have been calling it for a couple of hundred years – which Robert Prentner has defended, together with Donald Hoffman, Chetan Prakash, Manish Singh, and Chris Fields (also a contributor to this volume). Both panpsychists and interface theorists share something crucial in common: rather than starting with the physical world and trying to

---

<sup>5</sup> I prefer Mørch's (2018) and my own (2021) philosophical reworkings of IIT.

account for consciousness in terms of it, we start with consciousness and build up from there. I'm delighted Prentner is on board with the aspiration to build a post-Galilean science of consciousness. It's early days in building this new paradigm, and I'm glad different people are developing different non-reductionist options to see where they lead. Having said that, in my personal view, a panpsychist approach is more promising.

Prentner raises two criticisms of my panpsychist approach. Firstly, he claims that, in contrast to the interface theory, panpsychism remains a kind of dualism. Secondly, Prentner suggests that my approach in particular is too hostile to mathematics. The first criticism involves a misunderstanding of the theory. As I explain in my response to Carroll above, Russellian panpsychists don't think that particles have physical properties and consciousness properties; the claim is rather than the physical properties of the particles *are* forms of consciousness. To be fair to Prentner, he accurately reports that we Russellian panpsychists distinguish between the dispositional properties that physics tracks and the intrinsic nature that underlies them. But these aren't really two different kinds of property. All that really exists, according to Russellian panpsychism, are various forms of consciousness. The 'dispositional properties' are just a way of talking about what those forms of consciousness do. Matter is what consciousness does.

As for the accusation that Russellian panpsychists are hostile to mathematics, Prentner never substantiates this claim. The only claim we make about the limitations of mathematics is that you can't fully capture the qualities of consciousness in the purely quantitative vocabulary of mathematics. But Prentner agrees with that claim! It's true that there's not much mathematics in *Galileo's Error*. But that's because this is a book that aims at *laying the foundations* for a science of consciousness rather than itself engaging in first-order science. Panpsychists can employ mathematics as much as interface theorists, as can be seen by the highly mathematical work of Integration information theorists.

I turn now to concerns I have with the interface theory, as defended by Prentner, Hoffman, et al. These objections are intended in the spirit of collegial challenge. As I understand the interface theory, it commits to networks of conscious agents underlying the physical world. To make good on this commitment, interface theorists will have to postulate structures of conscious agents that facilitate not only the emergence of the structures of fundamental physics, but also the emergence of subjects like us who have the experience of interacting with a virtual world that can be accurately predicted by the equations of physics. This is a Herculean task. It's been hard enough to get the equations of physics we already have. Coming up with a whole new level of mathematical structure underneath that, which yields precisely the same predictions, is rather a large challenge. Moreover, I can't see the motivation for taking on that challenge. Rather than postulating a *new* level of mathematical structure *underneath* physics, why not locate the networks of conscious agents *inside* the mathematical structure of physics itself? This is essentially the Russellian panpsychist approach.

Taking the panpsychist approach also provides a wonderfully simple account of why it is that we emergent subjects experience a world that is accurately described by the equations of physics. According to Russellian panpsychism, there really is a world that is accurately described the equations of physics – networks of conscious agents realising the

mathematical structure of physics – and we perceive that world. In contrast, on the interface theory, we emergent subjects are interacting with networks of conscious subjects that are *not* accurately described by physics, in such a way that we seem to experience a world that *is* accurately described by physics. Dr Prentner (Hoffman, et al): cut out the middle man! Much simpler to suppose that we seem to experience the world physics describes because we *do* experience the world physics describes.

Donald Hoffman has argued that it is highly unlikely that we experience the world we seem to experience, given that our senses have evolved for survival fitness rather than for truth. This is supposedly proved by the ‘Fitness Beats Truth Theorem’ (Prakash, et al 2017). My objection to this argument – as I put to Hoffman during our three-hour discussion for Annaka Harris’ forthcoming audio series – is that it overgeneralises in a problematic way. If we ought to doubt the testimony of our senses, then we ought similarly to doubt the testimony of our evolved capacity for forming judgements concerning the mental states of others. We are hardwired to judge the emotions of others on the basis of their behaviour and facial expressions. But if this hardwired capacity was evolved for truth rather than survival, and if this is sufficient for us to reject the deliverances of our sensory perception, then we ought likewise to reject the deliverances of our judgements about other minds. We ought to think others are zombies, or at least have no faith in our judgments that crying indicates sadness. I consider this a *reductio ad absurdum* of the ‘Fitness Beats Truth’ argument.

#### Reply to Chris Fields

As with Prentner, I am inclined to think Chris Fields and I are basically on the same path. The goal for both of us is a robust, empirically adequate scientific theory in which the reality of consciousness is fundamental. This is what I mean by the ‘post-Galilean science of consciousness.’

However, like Prentner, Fields mistakenly takes me to be hostile to mathematics. But, again, my only claim is that the qualities of experience cannot be reductively accounted for in mathematical terms, and – like Prentner – I think Fields agrees with this. It’s true that I don’t employ much mathematics myself, but that’s because I’m a philosopher trying to lay the groundwork for a science of consciousness, rather than a scientist formulating a first-order theory.

I’ve suggested we’re basically on the same path, but Fields’ main line of critique involves a skepticism about some of my fundamental aims. Fields questions the need for a theory of qualia *per se*, i.e., a theory that explains, for any given experience, why the experience has the qualitative character it does; and he diagnoses my desire for a theory as arising from two sources: (a) an assumption that qualia are compositional, (b) a yearning for human specialness. I reject both of these diagnoses. Strong emergentist panpsychists may reject compositionality about qualia. And whilst Chapter 5 of *Galileo’s Error* explores the implications for human existence of the Russellian panpsychism that is defended in earlier chapters, the arguments in earlier chapters are certainly not rooted in a desire for a world consonant with human happiness. I’m interested not in the view that I’d like to be true but in the view that’s most likely to be true.

At least this is what I aspire to, and I believe I succeed fairly well at separating my theoretical work from my personal desires. As far as I can judge it, I really am obsessed with truth, with trying to have my best guess as to what reality is like. Maybe I'm totally deluded in this regard, and my fundamental motivation is yearning for the Cosmic Daddy of my Catholic upbringing. But any theorist could be psychoanalysed for the 'real' motivations underlying their apparent desire for truth. The Churchland/Dennett-esque idea that science as we currently conceive it is just obviously the path to all truth is terribly comforting.

My basic assumption here is that reality is intelligible. We have to start with some basic, unexplained facts, but there should be in principle an intelligible story as to how non-fundamental facts emerge from fundamental facts. Human qualia are real. And therefore we need an explanation of *what is going on in reality* to account for their existence. Perhaps, as the dualist or the emergentist panpsychist supposes, human qualia are fundamental features of reality. Perhaps, as the materialist or the reductive panpsychist supposes, human qualia can be reductively explained in terms of more fundamental features of reality (just because *some* consciousness is fundamental, it doesn't follow that *human* consciousness is fundamental). Either way, the fundamental drive of science and philosophy is to explain.

Fields wonders why it is 'satisfying' to postulate consciousness as the intrinsic nature of matter, again suspecting that I'm looking for a view that 'satisfies' in some emotional sense. I take it that what we are looking for is the most parsimonious account of the data. As explained above, Russellian panpsychism offers a radically non-dualistic, and therefore radically parsimonious, theory. Rather than postulating physical properties *and* consciousness properties, we identify the two. Physical properties just are consciousness properties characterised in terms of how they behave. Like most of the scientists in this volume, Fields misinterprets Russellian panpsychism as the view that consciousness, as the intrinsic nature of matter, doesn't do anything. Quite the contrary, if consciousness is all there is, either consciousness does something or nothing does anything.

## **Part II: Replies to Philosophers**

### Reply to Luke Roelofs

I don't have too much to say in response to Luke Roelofs excellent essay, simply because I am entirely in agreement with most its contents. Roelofs does an amazing job debunking the various charges that panpsychism is 'anti-science'.

Having said that, there are some slight differences between the two of us. If panpsychism is a middle way between dualism and materialism, I am slightly closer to the dualist pole and Roelofs to the materialist. Roelofs once told me that if they weren't a panpsychist they'd be a materialist, where I would certainly embrace dualism before materialism. As Roelofs says, I'm less convinced of micro-reductionism than they are. Roelofs says micro-reductionism fits better with the trajectory of science. But whether or not everything that happens is totally determined at the level of fundamental physics is an empirical question, and I don't think

we're yet in a position to judge empirically whether it's true (see my reply to Carroll). Until such a time as we can assess the view empirically, I'm not sure it's appropriate to draw inferences about the nature of reality from the hunches and methodological assumptions of working scientists.

Overall, I'm much more persuaded of the clash between materialism and consciousness realism than I am of the clash between dualism and the empirical facts. And given my greater affinity with dualism than materialism, I'm not totally ready to sign up to the 'Monist United Front' Luke proposes (opposition to dualism is a tentative third clause in the post-Galilean manifesto). Maybe if I set up the 'Anti-Materialist United Front' in the building next door, Roelofs and I can run a tea tent in the space between the two organisations.

### Reply to Annaka Harris

Annaka Harris' solution to the combination problem is intriguing and something I will continue to reflect on. I'm not yet convinced, however, that it resolves all of the challenges that go under that banner. Harris proposes that all physical systems are associated with some form of conscious experience, presumably both at the micro and the macro level. But if consciousness exists both at the neurophysiological level and at the level of fundamental physics, a crucial theoretical question arises:

*The Reduction Question* – Can the consciousness of the brain be reductively explained in terms of the consciousness of the particles making up the brain, in something like the way the liquidity of water can be reductively explained in terms of the underlying chemistry?

If we want to answer 'yes' to the reduction question, then the combination problem returns as the challenge of how to make sense of such a reduction. But if the answer is 'no', the spectre of epiphenomenalism looms again. If Roelofs (this volume) and Carroll (this volume) are right that all of my behaviour can be causally explained in terms of micro-level goings on that are the proper subject of matter of physics, then the consciousness that exists at the macro-level of the brain – what we think of 'Philip Goff's consciousness'<sup>6</sup> – (if irreducible) has nothing left to do, no role to play in governing behaviour. I know from her work more generally that Harris doesn't believe in free will, but it seems a more extreme step if the contents of my mind have absolutely no impact on the physical world, e.g. if the ideas I am thinking of in composing this reply play no role in determining what words I'm typing on the page.

In my recent work (Goff 2021), I have developed a form of 'hybrid panpsychism', which tries to address this dilemma by distinguishing sharply between subjects and their experiences, holding that the former are 'strongly emergent' (i.e. they can't be reductively explained) whilst the latter are 'weakly emergent' (i.e. they can be reductively explained, in terms of consciousness at the level of physics). My hope is that that this combination of strong and weak emergence can allow me to avoid *both* the combination problem (as I'm not trying to

---

<sup>6</sup> I am just pointing to a certain form of experience and do not mean to commit to a self/subject that has that experience.

reductively explain subjects of experience) *and* the threat of epiphenomenalism (as human experience can be reduced to, and thereby derives its causal power from, the level of fundamental physics). Of course, this strategy involves precisely the commitment to subjects that Harris rejects. But it seems to me that this allows me a way of avoiding problems that I can't see how Harris's form of panpsychism can avoid.<sup>7</sup>

### Reply to Damian Aleksiev

Whilst most discussion of panpsychism focuses on the combination problem, Damian Aleksiev is pioneering work on a quite different challenge: what he calls 'the missing entities problem.' I will here sketch some tentative responses to this new set of challenges, which I hope to give a fuller response to in future work. Aleksiev's concerns seem to be instances of the *structural mismatch problem*: the challenge for the panpsychist of explaining how the structure of human consciousness relates to the structure of consciousness at the fundamental level (which, for the panpsychist, is the level of basic physics). To take a very simple form of this: the physical structure of the parts of my brain which support my consciousness are highly complex, if we take it down to the detail of every individual quark and electron, whereas that enormous complexity doesn't seem to be present in *my* consciousness.

Wrestling with the structural mismatch problem has been my main motivation for returning to a 'cosmopsychist' form of panpsychism, according to which the fundamental forms of consciousness are the intrinsic natures of universe-wide fields, forms of consciousness which are borne by the universe as a whole. Some of that very complex experience borne by the universe is located in my head. But, on the hybrid cosmopsychism theory I have developed (Goff 2021), that's not *my* experience. Rather, my experience consists of a 'thinned-out' version of the very busy experience borne by the universe inside my head. To make this clearer, consider an analogy. Imagine someone experiencing a countryside scene of a lake, trees and birds. Now suppose we take out some of the detail, so that the person is only experiencing the lake without the trees and birds. The latter experience would be a 'thinned-out' version of the former. Similarly, my experience is a thinned-out version of the very rich experience inside my head: it's what results when you take the very busy experience inside my head (corresponding to the very complex physical structure at the fundamental level of the brain) and take out a lot of detail.

It's not clear to me why this same account could not be used to address at least the spacetime gap and the quantum gravity gap. Perhaps experience at the fundamental level is timeless and has an esoteric geometrical structure. But the process of thinning out that structure to form my experience will result in experience with a quite different structure. It may very well lack the metric of fundamental spacetime, and the very selective cutting and pasting involved in putting together my experience may be precisely what yields its temporal character.<sup>8</sup>

---

<sup>7</sup> I should say that Harris responded to Goff 2021 in an earlier version of her contribution, but I advised as editor focusing on laying out her own position, given the tight word limit.

<sup>8</sup> It's true that my most recent version of panpsychism is not fully the constitutive version Aleksiev is critiquing in his paper. However, although I have departed from fully constitutive panpsychism, I have done so in order to address the combination problem not the structural mismatch problem.

The quantum state gap is more troubling, as it's not clear to me that we could get from the structure of the quantum state to the structure of human consciousness via thinning out. To make this clear consider, for the sake of simplicity, a situation we never actually find in the actual world. Suppose all of the particles in the universe were in an utterly determinate state. This corresponds to all the amplitude of the wave function occupying a single location in  $3 \times N$  space. Interpreted as a form of panpsychism, that amplitude is a form of consciousness. But it's just a blob lacking any kind of spatial extension or internal structure. It's clear that we can't get from a simple structureless blob to the complexities of human consciousness by taking away some detail.

The single most important claim of *Galileo's Error* is that we need to take the reality of consciousness, as we know it 'from the inside', as a fundamental scientific datum in its own right. If that datum rules out certain interpretations of quantum physics, e.g. those in which fundamental reality is entirely constituted by the wave function, then we should not hesitate to reject those interpretations on that basis. Indeed, this could be one important way of deciding between the many interpretations of quantum mechanics, and an important aspect of the post-Galilean research programme (cf. Cortes, Smolin and Verde's contribution to this volume). I have explored this issue in detail in my paper 'Quantum mechanics and the consciousness constraint.'

#### Reply to Alyssa Ney

Alyssa Ney raises some great challenges to my views. In responding, I'd like to start with one small but important point which I think is a misunderstanding of my view. My view is not that we cannot use the tools of physical science to investigate consciousness, but rather that the tools of physical science *alone* cannot give a full account of consciousness. Neuroscience can provide us with an account of the physical correlates of conscious experience, but cannot, in my view, give a fully satisfactory explanation of *why* conscious experience is correlated with brain activity. I'm proposing *adding* to the toolbox of the science consciousness, not taking away from it.

Ney draws attention to the ambiguity in *Galileo's Error* regarding whether I'm subscribing to scientific instrumentalism or structural realism. This is partly the result of the necessity of being less precise in a book aimed at a general audience. It's much clearer (I hope!) in my academic book *Consciousness and Fundamental Reality* that I advocate structural realism. I appreciate that the line in *Galileo's Error* Alyssa points to – '..physical science doesn't even tell us what matter does...' – might seem to suggest otherwise. But whilst I think physical science doesn't strictly speaking tell us what anything does – that would require a non-circular specification of the manifestation of some physical property – I do think it can accurately capture the abstract causal structure of reality.

---

The basic idea I now refer to as 'thinning out' was present in the version of constitutive panpsychism I defended in *Consciousness and Fundamental Reality*. Hence, I'm not persuaded that the issues Aleksiev raises here give us grounds for dropping constitutive panpsychism.

One might get the impression reading Ney's article that I'm wildly at odds with orthodoxy in philosophy of science. However, as Ney says, structural realism is orthodoxy in philosophy of science, and to that extent I am completely in line with orthodoxy. It is true, as Ney says, that most philosophers of science would go for *ontic* structural realism – all that exists is structure – rather than *epistemic* structural realism – all science can tell us about is structure. But that's because most philosophers of science aren't thinking about how to accommodate consciousness into their view, and to the extent that they do, they're likely to adopt materialism. I don't think I've kept hidden that I'm proposing radical change – the subtitle of *Galileo's Error* calls for a 'New Science of Consciousness' – and to that extent I'm challenging orthodoxy. Whether that radical change is called for depends on whether the arguments against materialism about consciousness are successful (I briefly outline these arguments above in my replies to Rovelli and Seth), and Ney doesn't in this article deal with those arguments.

On free will, I agree with Ney that a compatibilist would be unmotivated to consider the 'pan-agentalist' position I explore in the final chapter. My argument here is conditional: *if* you're a libertarian about free will, you should be a pan-agentalist; but I don't offer an argument for the antecedent of that conditional. Again, this might not have been completely clear due to the necessity in a popular book of not defining everything totally precisely; the argument is (I hope!) laid out in a more rigorous fashion in my article 'Panpsychism and free will.'

On the foundations of morality, I agree that without some way of accounting for prudential oughts, the account I sketch wouldn't provide an objective basis for morality. But it may perhaps reveal some false presuppositions in egoism; and if you combine the view I outline with internalism about reasons – roughly, the view that one has reason to do what one desires – then we may have an objective basis for demonstrating that those with self-interested desires ought to have altruistic desires.

I also agree with Ney that if analytic naturalism were plausible, Sam Harris would have a way of avoiding the concerns I raise with his view; and perhaps that is Harris's view. Having said that, I don't think analytic naturalism is a very plausible view. For one thing both ethical egoism – the view that I only have reason to do what is good for me – and non-egoistic ethical realism – the view that I have reason to do both what is good for others as well as myself – seem quite clearly non-contradictory, which leaves analytic naturalism with no resources to say which of these mutually contradictory views is correct.

Ney points out that there are other options for naturalistic reductions of ethics, which is of course correct. For what it's worth, I favour pretty hardcore non-reductionist realism about value, but I certainly don't take myself to have defended that in *Galileo's Error*. With both the discussions of free will and morality, there are a hundred different positions, and if I'd spent time precisely articulating what I took myself to have shown, then this chapter would have been a lot less accessible. One useful thing about this volume is that I can further clarify some of my views, and I'm grateful to Ney for pushing me to do so.

On the big picture meaning of life stuff, my core concern with materialism is that it is incompatible with the reality of conscious experience, and hence our official worldview is

incompatible with the thing of which we are most certain and the thing which gives life meaning and value. I appreciate that what I've just said assumes the falsity of materialism, which Ney disputes. At the end of the day, everything will hang and fall on the arguments against materialism about consciousness.

### Reply to Keith Frankish

Is Galileo's Real Error believing that qualities exist anywhere? Reading Keith Frankish's ingenious argument, I can almost go along with him. I have learnt so much from thinking my way inside Frankish's worldview, occupying his world for thirty minutes or so (no longer, in case I never get back...). But I can't help thinking it doesn't quite work out.

Let's suppose that Frankish is right that the sensory qualities – colour, sounds, smells, tastes – don't exist anywhere, either in the mind or the external world. Focusing on the case of colour, David Chalmers uses the term 'Edenic' for colours, sounds, smells, tastes, etc. as we naively take them to be – sensory qualities as they were in the Garden of Eden. We can thus define Frankish's 'Edenic illusionism' as follows:

1. The mind represents physical objects to have Edenic qualities.
2. Edenic qualities don't really exist, either in the mind or the external world.

It's clear that the mind can represent things that don't exist, e.g. if I hallucinate a pink elephant. Similarly, just because the mind represents objects to have Edenic qualities, it doesn't mean that they really do.

Does the hard problem of consciousness go away if we accept Edenic illusionism? My case against materialism is built upon the claim that we can't account for the qualities of consciousness in the purely quantitative language of physical science. But if these qualities don't exist at all – either in the mind or the external world – then haven't we entirely removed all the troubling phenomena that threaten the materialist worldview?

I don't think so. Even if my mind doesn't contain Edenic qualities, it nonetheless has the property of *representing* Edenic qualities. And the property of representing Edenic qualities is just as irreducible as Edenic qualities themselves. The argument against materialist reduction of Edenic qualities is roughly as follows (see my reply to Rovelli above):

1. Edenic qualities can be reductively explained in the terms of physical science only if Edenic qualities can be fully captured in the terms of physical science.
2. Edenic qualities cannot be fully captured in the terms of physical science.
3. Therefore, Edenic qualities cannot be reductively explained in the terms of physical science.

The same form of argument applies to the property of representing Edenic qualities:

1. The property of representing Edenic qualities can be reductively explained in the terms of physical science only if the property of representing Edenic qualities can be fully captured in the terms of physical science.

2. The property of representing Edenic qualities cannot be fully captured in the terms of physical science.
3. Therefore, the property of representing Edenic qualities cannot be reductively explained in the terms of physical science.

In order to fully describe the property of representing Edenic red, you'd have to fully describe Edenic red. Thus, if you can't do the latter in the purely quantitative vocabulary of the physical science, then you can't do the former in the purely quantitative vocabulary of physical science.

I often speak of the core challenge for the materialism as that of accounting for the 'qualities' of consciousness. And I am indeed inclined to the Galilean view that the sensory qualities that seem to be in the external world are really in the mind (see my reply to Liu below). But even if, strictly speaking, there are no qualities in consciousness, we still face the challenge of accounting for the property of consciously representing qualities. Either way, I don't think Frankish's strategy can save materialism.

#### Reply to Michelle Liu

Michelle Liu argues makes a powerful case for the for the polar opposite position to Frankish, for the real existence of Edenic qualities. Colours as we naively take them to be really do exist out there on the surfaces of objects. Call this view 'Edenic realism.'

The first thing I would note is that the existence of Edenic properties is subject to doubt in the way that the qualities I directly apprehend in my experience are not.<sup>9</sup> I can doubt that the rose in front of me is really red – I can even doubt whether it exists – but I cannot doubt that my experience of the rose has a reddish qualitative character

Moreover, I am not persuaded by Liu's theoretical argument for Edenic properties: that if we deny their existence, it is extremely hard to explain how we come to perceptually represent such non-existent properties. As David Hume observed, we 'gild and stain' the world with our sentiments. When I see something nauseating, I experience the 'disgustingness' as an intrinsic property of the object of my disgust, even though, of course, there isn't really any such intrinsic property as 'disgustingness'. I don't see why something similar couldn't be going on in the case of colour experience: I misrepresent an intrinsic quality of my experience of a rose as an intrinsic quality of the rose itself. I agree with Liu that the experience of colour is much more robust than the experience of disgustingness; the former as opposed to the latter is not dependent on past experience and is 'just as perceptually salient and persistent as that of shape and size.' But that is presumably because the way we experience colour is hardwired into us from birth, whereas the experience of disgustingness is partly shaped by our life experience. This difference seems to me perfectly compatible with the commonality that they both involve misrepresenting an intrinsic quality of experience as an intrinsic quality of the object of experience.

---

<sup>9</sup> Or at least the fact that I consciously represent those qualities is not subject to doubt (see my reply to Frankish).

I'm more persuaded by Liu's argument that a world without Edenic properties is of significantly less value than the naïve world, due to the absence of the beautiful colours that seem to fill our world. This is a consideration I hadn't properly appreciated before. Of course, one could retort 'Why think reality is likely to be of value?' But any kind of theorising about reality depends on anti-skeptical assumptions, e.g. that we're not in the Matrix, that other people have consciousness. Perhaps Liu can make a case that assuming Edenic realism is justified by a similar kind of anti-skeptical assumption.

However, ultimately I think there are more powerful considerations on the other side of the ledger. My main motivation for rejecting Edenic realism is due to issues in the philosophy of perception, which aren't Liu's focus here. Edenic realism goes naturally with a 'naïve realist' view of perception, according to which perception involves a *direct* connection with things in the world: when I'm looking at a rose, the red quality in my experience just is the red quality of the rose. Worries arise when we try to square this with the existence of hallucinations. Clearly when I'm hallucinating a red rose, I'm not in direct contact with the Edenic redness of a physical object, as there's no red object there. In this case, then, the redness must be in my head. So it seems that we must say that when I'm veridically perceiving a red rose, what I'm in contact with is redness out there in the world, whereas when I'm hallucinating a red rose, the redness is in my head.

I think there are a number of problems with this position, which I hope to develop in future work. Just to focus on one issue: the mental redness in the hallucination has the same character as the Edenic redness in the veridical experience. How does the cognitive system manage to locate a mental property to use in the hallucination with the same character as the Edenic property of the veridical case? And why have we evolved to create a deceptive experience in hallucination? If the cognitive system somehow 'knows' it's hallucinating (which it presumably must do as it activates the mental redness when and only when there is a hallucination of red) why not flag up that this is a hallucination rather than encouraging the person to think it's a true experience of red by activating a mental quality that resembles Edenic redness. If these issues could be addressed, I would be willing to look again at Edenic realism, but for the moment I remain unpersuaded.

#### Reply to Alex Moran

Like Liu, Alex Moran is inclined to believe in Edenic qualities, and hence worries that my panpsychist view is at best an incomplete theory of reality: even if it can account for the qualities of the mind, it is unable to account for the qualities of the external world: the colours, sounds, smells and tastes. He then offers a compelling presentation of two alternative views with the potential to do both. My reply to Liu above, in which I have argued against the existence of Edenic properties, doubles as a reply to Moran: if there are no Edenic properties, then there is no additional explanatory obligation for the panpsychist over and above accounting for the qualities of the mind. I think Moran is right that other forms of Russellian monism work better if we're trying to account for both experiential qualities and Edenic properties. But, given my doubts about Edenic properties, I'm not yet motivated to embrace these alternatives to panpsychism.

The other fascinating aspect of Moran's proposal is his articulation and defence of *grounding physicalism*, which is an emerging position in the philosophical literature more generally. I certainly agree with Moran that there should be room for a physicalist position according to which the relationship between mental properties and physical properties is looser than identity (in *Galileo's Error*, but not in my academic work, I deliberately oversimplify by construing materialism as an identity theory). Philosophers tend to call this looser relationship 'grounding' whilst scientists call it 'emergence.' However, I struggle to make sense of a grounding relation that involves no explanatory connection between the fundamental and the non-fundamental. Surely physicalists must *explain* the existence of consciousness in terms of the physical properties of the brain, in such a way that the explanatory gap is closed.

Moran resists this by comparing grounding to causation, which many philosophers now take to be compatible with the existence of explanatory gaps: what causes what is discovered through empirical investigation not rational deduction. But, in that case, I don't understand the difference between 'grounding' and 'causation.' David Chalmers is associated with the property dualist position according to which consciousness is causally brought into being by the physical properties of the brain, in accordance with fundamental psycho-physical laws of nature. I can't see the substantial difference between that position and the position Moran and others are calling 'grounding physicalism.'

#### Reply to Ralph Weir

Ralph Weir presents a powerful argument that the post-Galilean science of consciousness should embrace substance dualism. I think he's right one can't accept the logical possibility of a zombie (a body without a mind) whilst denying the logical possibility of a ghost (a mind without a body). However, the real-world implications of a possible ghost are less straightforward than the real-world implications of a possible zombie. Crucially, I'm not persuaded that the mere possibility of my ghost twin entails that *I* could exist in a disembodied state.

I think the issue hangs and falls on the metaphysics of substance. Roughly speaking, there are two rival views: bundle theory and substance-attribute theory. According to bundle theory, a physical object is just a bundle of its properties: an electron, for example, is nothing more than its mass, its negative charge, and so on, 'bundled together' (however that is cashed out). According to substance-attribute theory, in contrast, there aren't merely the properties of the electron, but also the thing that *has* the property, also known as the electron's 'substratum.'

I think Weir's argument is sound if we assume bundle theory. If my conscious mind is nothing more than a bundle of its conscious states, and that bundle of conscious states could exist in the absence of anything else, it follows that my conscious mind could exist in the absence of anything else. However, now suppose my conscious mind is a substratum that *has* its conscious states. Is that substratum *essentially* tied to any its properties? Four possibilities suggest themselves:

1. **My substratum is essentially tied to experiential properties and only to experiential properties** – My substratum can exist only in so far as it is bearing experiential properties, but it could exist without bearing any other kinds of property.
2. **My substratum is essentially to experiential properties but also to certain physical properties (for the purposes of this discussion let's suppose – contrary to panpsychism – that physical properties are wholly non-experiential)** – My substratum can exist only in so far as it is bearing experiential properties and certain physical properties, but could (and perhaps does) exist without any other forms of being.
3. **My substratum is essentially tied to certain physical properties but not to experiential properties** – My substratum could exist without bearing experiential properties but could not exist without bearing certain physical properties.
4. **My substratum is not essentially tied to any of its properties** – On this view, I could exist as pretty much anything: an abstract object, a non-experiencing boiled egg, whatever.

I don't think we can settle on introspective grounds which of these options is correct. Through introspection, I know about the essential nature of my conscious states, and I know that *those states* could exist without any other properties. I arguably also know that there is something that bears my conscious states: my substratum. But I have no way of knowing on grounds of introspection whether, in conceiving of my substratum as *a bearer of conscious states*, I thereby grasp its complete, essential nature. This seems me essentially the flaw in Descartes' argument for dualism, and the worry remains, I think, for Weir's argument.

To be fair, Weir does acknowledge something like this response, referring to it as belief in 'transcendent egos,' saying in a footnote that these might be conceived of as substrata. However, that might suggest a certain 'add on' to a theory of mind. However, the choice between bundle theory and substance-attribute theory is a fundamental theoretical choice-point we face in deciding, not merely on a theory of mind, but on a theory of reality itself. Which way we go with regards to that fundamental choice-point will determine whether or not Weir's argument is sound, or so it seems to me.

### Reply to Galen Strawson

Galen Strawson and I have the same view about the nature of reality. We both believe that consciousness can't be accounted for in the terms of physical science, but that that's okay because physical science doesn't tell us the nature of matter. Consciousness is totally physical, but it's an aspect of the physical that goes beyond what physical science can teach us about. We're also both inclined to think that consciousness exhausts the nature of physical reality. In terms of living philosophers, Strawson is the father of the post-Galilean revolution.<sup>10</sup>

---

<sup>10</sup> I will always be deeply indebted to Galen for inspiring me with the confidence to defend a whacky view for no other reason than that it's probably true.

Despite our agreement on issues of substance, we've had a fifteen-year long dispute about what to call the view we both accept. Strawson is very passionate about calling it 'physicalism' and/or 'materialism'. I think it's better to use those words to describe the view we're opposed to, namely that the nature of consciousness can be entirely accounted for in the terms of physical science.

It doesn't seem to me that there's any fact of the matter we're arguing over here. As Humpty Dumpty wisely observed, we can use words how we wish. Strawson is better versed than I in historical matters, so I won't dispute his claim to be using the word 'materialism' in a way that fits better with its historical usage. However, I think we should decide how to use the words 'materialism' and 'physicalism' on pragmatic grounds. Specifically, I think we should go with the meanings of these terms that are most useful for having the debates we want to have. There's a big and important debate between those who do and those who don't think consciousness can be accounted for in the terms of physical science, and to have that debate, it's incredibly useful to have words that identify people in terms of which side of that debate they're on. We can't say 'dualists' and 'anti-dualists', as people on our side of the debate include panpsychists and neutral monists. I think it makes sense to use the terms 'physicalists/materialists' and 'anti-physicalists/materialists.'

Part of what's going on here is that Strawson has very little time for the view of our opponents – the group I call 'physicalists' and he calls 'physicSalists.' Strawson (2018) argues that physicSalism amounts to the denial of the reality of consciousness, which he has famously called 'the Denial' and the 'silliest claim ever made.' Of course, I agree with Strawson that we can't fully account for consciousness in the terms of physical science, and to that extent I agree that physicSalism does commit one to the denial of the reality of consciousness. Where I think we disagree is that I don't think that deep philosophical truth is *obvious*, such that anybody who denies it is in denial or deluded. Many of the physicSalists in this volume – Ney, Carroll, Rovelli, Seth – fully believe in the reality of consciousness – feelings, experiences, sensations – whilst also believing that physical science can in principle give a full explanation of consciousness. I think they're wrong but I don't think they're stupid for thinking this. Philosophy is hard!

Having said, there are the physicSalists who explicitly *do* deny the reality of phenomenal consciousness, such as contributor to this volume Keith Frankish. I agree with Strawson that this view is very implausible. But, again, I think philosophy is hard, and I can understand how honest and earnest philosophical reasoning can get to you to that position. Of course, in some sense I think that reasoning is flawed; that follows from the fact that I reject that philosophical view. But there's plenty of honest and earnest disagreement in philosophy.

Moreover, I think Strawson and I should welcome the growth of illusionism about consciousness. The more it becomes accepted that the choice is between illusionism and anti-physicSalism, the more I believe society will come to reject the physicSalism we both find so implausible. Indeed, the growth of Dennett/Frankish-style illusionism as the dominant materialist position may ultimately be the sociological cause of the post-Galilean science of consciousness moving from fringe to mainstream.

### Part III: Replies to Theologians

#### Reply to Joanna Leidenhag

Joanna Leidenhag does a great job of describing more nuanced ways of understanding the relationship between God and the world. But I think maybe she's reading a little too much into my references to God in *Galileo's Error*. These discussions weren't really intended to engage with the question of miracles or God's existence, but rather to illustrate the challenges to dualism, which I tried to illustrate by imagining a God who intervenes *very regularly*, and hence makes Her presence very well known. As I say in the book, '...if God does intervene in the world, She doesn't act enough to make Her presence *obvious*' (Goff 2019: 37). My fundamental objection to traditional theism is not to do with the possibility of miracles but is rather rooted in the traditional problem of evil and suffering. I think it's highly implausible that a loving God who could do anything would create and sustain a world with so much pain. Obviously, this is a huge debate I can't get properly into here.

My disagreements with Leidenhag begin in the latter half of the article, where she presents an intriguing argument that the case for panpsychism should also lead one to theism. Galen Strawson (another contributor to this volume) argues for panpsychism on the grounds that we can't intelligibly get consciousness from non-consciousness, but I've never supported that argument. Whilst I don't think we can intelligibly get from the purely *quantitative* facts of physical science to the *qualitative* facts of conscious experience, I'm not sure we can rule out that there are some special kinds of properties of matter which are not themselves forms of experience, but which somehow combine in various ways to produce experience; philosophers call these 'proto-phenomenal properties' (Goff 2006, 2009).

My argument for panpsychism is rather based on considerations of simplicity. I believe that we *can* account for human consciousness in terms of more fundamental forms of consciousness, and that such a panpsychist theory is the most parsimonious theory alternative to materialism (which I think should be rejected on the grounds that it can't account for consciousness). Panpsychism is the most parsimonious theory able to account for both the reality of consciousness and the data of public observation.

Leidenhag is correct that my case for panpsychism relies on the idea that there must be an intelligible connection between fundamental and non-fundamental facts. In my academic work, I call (a more detailed version of) this principle 'Minimal Rationalism.' We could roughly define it as follows:

*Minimal rationalism (MR for short)* – For any non-fundamental fact F, it must be possible in principle to derive F from the complete fundamental story of reality.

However, I think MR is a much more modest principle than the Principle of Sufficient Reason (PSR). It's one thing to say that there must be an intelligible connection between the fundamental and non-fundamental facts, quite another to say that there must be an intelligible story about why the fundamental facts exist in the first place. Therefore, I think the panpsychist can quite consistently defend her position whilst rejecting theism by defending MR whilst rejecting PSR.

Having said that, I am rather sympathetic to PSR. I'll say more on that presently, but let's first think about where the PSR leads us. An infinite regress of universes is no good, as then you don't have an explanation of why the infinite regress of universes has always existed (as opposed to, say, an infinite regress of ghosts, or nothing at all). The only way there could be a complete explanation of everything is if there is something that explains its own existence; PSR leads inexorably to the postulation of a *self-explainer*.

How can we get a grip on the idea of a self-explainer? Perhaps by considering its polar opposite: an impossible being. An impossible being is a being whose nature – or rather the nature it would have if it existed – explains its non-existence. We have a good grip on many impossible beings, for example, square circles. Once you grasp the nature of a square circle, you just see that it can't possibly exist. By analogy, a self-explainer would be a being such that, once you grasped its nature, you'd see that it *must* exist. We have no positive understanding of such a being, but that doesn't mean there isn't such a thing. Dogs can't do mathematics; maybe forming a positive conception of a self-explainer is similarly beyond us. The limits of human comprehension are not the limits of reality.

If we accept PSR, we have to postulate a self-explainer. But why accept PSR? Quite simply, it seems to me that, all things being equal, a theory of reality in which there are no brute facts is superior to a theory of reality in which there are brute facts. This principle leads us to postulate a self-explainer – in order to avoid brute facts – unless there is some comparable theoretical reason to avoid postulating a self-explainer.

Could Occam's razor give us grounds for avoiding postulating a self-explainer? Not if we think of Occam's razor, as I do, as a principle telling us what kinds of explanation we should seek – the most parsimonious ones – rather than a principle telling us what should or shouldn't be explained in the first place. I suppose someone might say that we shouldn't postulate something that does explanatory work when we don't understand how that explanatory work is done. But that seems to me false: we can have very good reason to suppose that something had a cause, even if we don't understand how that cause did its causing. Thus, I can't see any good reason *not* to postulate a self-explainer, and in the absence of some such reason, the theoretical attraction of avoiding brute facts ought to lead us to postulate a self-explainer.

For these reasons, I think there probably is a self-explainer. However, I don't see why the self-explainer can't be the universe itself. Leidenhag raises three objections to this view, which I lay out below together with my responses:

1. Objection: The universe began to exist, and hence doesn't exist necessarily.  
Response: It's true that the *temporal phase* of the universe began to exist, but it could be that, in the absence of time, the universe exists in a timeless form. Leidenhag may object that we have no reason to think the universe could exist in a timeless form. I would respond: PSR entails we must choose between two hypotheses: a self-explaining universe which can exist in a timeless form, or a supernatural self-explainer distinct from the physical universe. The former hypothesis is more parsimonious, and therefore the one we ought to go for.

2. Objection: If the universe exists necessarily, nothing is contingent.  
Response: I don't think that's quite right. If there is quantum indeterminacy, then this will introduce an element of contingency. And the necessity of the universe's existence is consistent with the emergence of libertarian free will.
3. Objection: It is conceivable that the universe doesn't exist, and my commitment to the zombie argument involves a commitment to conceivability entailing possibility. Putting these together, I ought to believe in the possible non-existence of the universe.  
Response: It's also conceivable that God doesn't exist. Or at least it's conceivable that an all-knowing, all-powerful and perfectly good being doesn't exist. Of course, we define 'God' as existing necessarily, but I could equally define a new term 'super-universe,' which refers to the universe but stipulates that it exists necessarily. I think that if there is a God who exists necessarily, it must be in virtue of some aspect of Her nature beyond our comprehension. Why not take the more parsimonious option of ascribing that nature to the universe itself?

### Reply to Sarah Lane Ritchie

Sarah Lane Ritchie is here arguing for a conclusion that is often dismissed as 'new age fluffy thinking'. But she argues with such rigour, clarity and nuance, that such an accusation would be ridiculous, rooted only in ignorance and prejudice.

It's important to note that many philosophers and scientists defending panpsychism are completely atheistic secularists. Luke Roelofs (a contributor to this volume) and David Chalmers are cases in point. They don't believe in a transcendent reality, but they do believe in feelings and experiences, and hence want to an account of how those things fit in to our overall theory of reality. I'm sure Ritchie would agree with all of this.

Having said that, I would say that panpsychism removes some reasons *not to believe* that mystical experiences are veridical. Suppose one has a mystical experience with the following content: *there is a higher form of consciousness underlying all things*. If you're a materialist, then you have to think this experience is a delusion, as what this theory suggests concerning fundamental reality is inconsistent with the story about fundamental reality we get from physics. But if you *already* think that fundamental reality is constituted of forms of consciousness, then it's not too much of a leap to suppose that the higher form of consciousness you seem to be aware of in having the mystical experience is also part of the fundamental story of reality.

But just because you have no reason *not* to believe X doesn't mean you have a reason *to believe* X. Richard Dawkins doesn't buy the problem of evil as a reason to doubt the existence of God, but he nonetheless ascribes a very small probability to God's existence because he thinks there's no evidence for it. Many people think there's no rational basis for the mystic to trust what her experience seems to be telling her about the nature of reality, as it could simply be a delusion caused by unusual brain activity. However, I'm sympathetic to William James' line (James 1902/1985: Ch. XVII) that this common viewpoint betrays a double standard. It's true that the mystic can't prove that their experience corresponds to an external reality. But nor can any of us prove that our sensory experiences correspond to

an external reality. All empirical knowledge begins with trusting what experience seems to be telling us. It's not clear how one could consistently tell the mystic it is irrational for her to trust her mystical experience without also denying the rationality of trusting one's own sensory experiences.

I've never had a mystical experience. But in my some of my deepest experiences – certain moral experiences, engagements with nature, deep meditation, or watching the light of early morning or late evening – I have a fleeting sense of a greater reality at the root of things. I wouldn't go so far as to say I *believe* that those experience correspond to an external reality. But I engage them, I trust them, I work with them in my spiritual practice. Maybe I'm a fool for taking these epistemological risks. But as John Locke wisely observed, 'He that in the ordinary affairs of life, would admit of nothing but direct plain demonstration, would be sure of nothing in this world but of perishing quickly' (Locke 1689/2008. Book IV, section 10).

## References

Boly, M., Massimini, M. Tsuchiya, N. 5 6, Postle, B. R., Koch, C., Tononi, G. (2017) 'Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence,' *Journal of Neuroscience* 4 October 2017, 37 (40) 9603-9613.

Buyse, Filip (2015). "The Distinction between Primary Properties and Secondary Qualities in Galileo Galilei's Natural Philosophy.' *Cahiers du Séminaire québécois en philosophie moderne/Working Papers of the Quebec Seminar in Early Modern Philosophy* 1: 20-43.

Chalmers, D. J. (1995) 'Facing up to the problem of consciousness,' *Journal of Consciousness Studies* 2: 3, 200-19.

Cobb, M. (2020) *The Idea of the Brain*, Profile Books.

Galilei, G. (1623/2008) *The Assayer*, in *The Essential Galileo*, trans. by Maurice A. Finocchiaro, 179-189. Indianapolis and Cambridge: Hackett Publishing Company.

Goff, P. (2006) 'Experiences Don't Sum,' *Journal of Consciousness Studies* 13: 10-11, 53-61.

Goff, P. (2009) 'Why Panpsychism Doesn't Help Us Explain Consciousness,' *Dialectica* 63: 3, 289-311.

Goff, P. (2017) *Consciousness and Fundamental Reality*, Oxford University Press.

Goff, P. (2019) *Galileo's Error: Foundations for a New Science of Consciousness*, Pantheon/Rider.

Goff, P. (2020) 'Panpsychism and Free Will: A Case Study in Liberal Naturalism,' *Proceedings of the Aristotelian Society* 120: 2, 123-44.

- Goff, P. (2021) 'How exactly does panpsychism explain consciousness?', available <https://www.philipgoffphilosophy.com/>
- Goff, P. (forthcoming) 'Quantum mechanics and the consciousness constraint,' in S. Gao (Ed.) *Quantum Mechanics and Consciousness*, Oxford University Press.
- James, W. (1902/1985) *The Varieties of Religious Experience, The Works of William James*, Vol. 13., Harvard University Press.
- Locke, J. (1689/2008) *An Essay Concerning Human Understanding*, edited by Pauline Phemister, Oxford University Press.
- Michel, M. (2019) 'Consciousness Science Underdetermined: A Short History of Endless Debates,' *Ergo*, 6: 28.
- Mitchell, Kevin (2019) 'Beyond reductionism – systems biology gets dynamic,' in *Wiring the Brain* blog, September 14, <http://www.wiringthebrain.com/2019/09/beyond-reductionism-systems-biology.html>.
- Mørch, H. H. (2018) 'Is the Integrated Information Theory of Consciousness Compatible with Russellian Panpsychism?' *Erkenntnis* 84: 5, 1065-1085.
- Nikos K. Logothetis (2008) 'What we can do and what we cannot do with fMRI,' *Nature* 453: 869–78.
- Odegaard, B. Knight, R. T. & Lau, H. (2017) 'Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception?' *Journal of Neuroscience* 4, 37: 40, 9593-9602.
- Papineau, D. (2001) 'The rise of physicalism,' in C. Gillett & B. M. Loewer (eds.) *Physicalism and its Discontents*, Cambridge University Press.
- Prakash C., Stephens K., Hoffman D.D., Singh M, Fields C. (2017) 'Fitness Beats Truth in the Evolution of Perception,' *Acta Biotheoretica*.
- Strawson, Galen (2018) 'The Consciousness Deniers,' *The New York Review*, March 13 <https://www.nybooks.com/daily/2018/03/13/the-consciousness-deniers/>