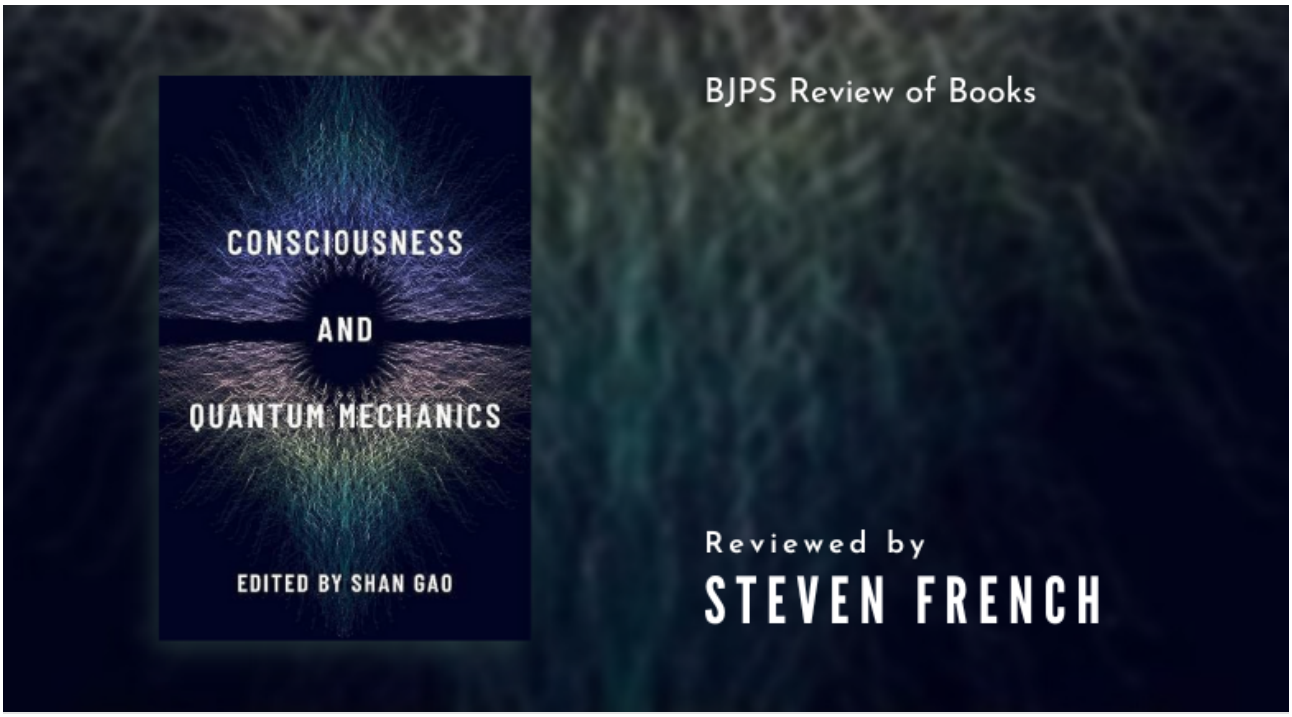

CONSCIOUSNESS AND QUANTUM MECHANICS

Shan Gao

Reviewed by Steven French

Consciousness and Quantum Mechanics
Shan Gao (*ed.*)
Oxford: Oxford University Press, 2022, £64.00
ISBN 9780197501665

For many years any mention of consciousness in the context of quantum physics was generally restricted to those popular accounts that might be found on the 'New Age' or 'Spiritual' bookshelves. Certainly, in 'mainstream' philosophy of physics, the concept was regarded as definitely *non grata*, following Putnam's ([1961]) and Shimony's ([1963]) famous set of critiques of the 'consciousness causes collapse' solution to the measurement problem in the early 1960s. Recently, however, consciousness has begun to tiptoe back into the limelight, as both explanans and explanandum. Here Shan Gao has collected seventeen contributions from prominent philosophers and physicists (including one Nobel Prize winner), which offer a disparate set of accounts of the role it might play. Following a helpful introductory orientation, these essays are grouped into three sections: 'Consciousness and Wave Function Collapse', 'Consciousness in Quantum Theories', and 'Quantum Approaches to Consciousness', although there is a certain degree of arbitrariness in the placement of some of the papers both within and between these divisions.

For example, Gao's own contribution, 'Why Mind Matters in Quantum Mechanics', appears mid-way through the collection, in the second section, but it would also serve as a useful entry-point—not least because most readers, I suspect, when they think about consciousness in the quantum context at all, have the measurement problem in mind. This is presented in mentalistic terms as involving the incompatibility between the following three assertions (p. 178):

(A1) The mental state of an observer supervenes on her wave function;
(A2) the wave function always evolves in accord with a linear dynamical equation, such as the Schrödinger equation;
(A3) a measurement by an observer yields a single mental state with a definite record.

Here Gao defends this formulation against recent criticisms, principally on the grounds that unlike the physicalistic version advocated by the likes of Maudlin ([1995]), for example, his is robust through changes in the overall interpretation of the theory. As he notes, this not only highlights the important role of the psycho-physical connection in generating the problem but offers a new perspective from which to view any putative solutions (pp. 180–81).

Of course, there is a significant question here: what is it about that psychophysical connection that prevents such mental states from themselves entering into a superposition, given that we do not appear to experience that? This is answered by Chalmers and McQueen in the opening essay, 'Consciousness and the Collapse of the Wave Function', by appealing to a rule that would prevent any superpositions from forming. The problem is, under such a restriction, the system would remain locked forever in a particular eigenstate of the relevant observable; and so, if consciousness or its physical correlate were that observable, we would never wake up from a nap (p. 27)!

To get around this, Chalmers and McQueen draw on a variant of the spontaneous collapse interpretation of quantum mechanics in which the collapse takes place gradually. Coupling this with a theory of consciousness known as 'integrated information theory', they argue that phenomenal 'qualia-shapes' may be approximately resistant through such a collapse, yielding determinate mental states (pp. 38–41).

Unfortunately, however, as Chalmers and McQueen's acknowledge, models in which the collapse proceeds slowly are difficult to reconcile with introspection, whereas their faster brethren may be ruled out by technologically feasible quantum computers (p. 44). More seriously, these approximate resistance models still allow for superpositions and when it comes to what it would even be like to be in such a state of consciousness, all that Chalmers and McQueen can do is gesture at the idea that it would involve 'some novel phenomenal mode of combination' of the relevant mental states (p. 52). For all that their discussion may serve as an 'existence proof for a relatively precise consciousness-collapse model' (p. 55), this objection remains the biggest obstacle that any such model must face.

In the following chapter, Okon and Sebastián claim to have overcome this obstacle with their 'subjective–objective collapse model'. They too associate a 'collapse operator' with consciousness and likewise appeal to a continuous spontaneous localization process. However, they then invoke the distinction between phenomenal and 'access' consciousness—where the former involves the experience we have and the latter what we come to be aware of—to explain why we fail to notice the transitions associated with a collapse (pp. 69–70).

A major problem with these kinds of approaches, however, is that they inherit all the well-known concerns associated with spontaneous collapse interpretations more generally: the strength of the collapse must be fixed in a somewhat *ad hoc* manner by a parameter that must not be too large, lest the quantum Zeno effect comes into play, nor too small, else macroscopic superpositions will be observed. And even setting that to one side, there remains the so-called tails problem, in that we never get a 'complete' collapse to a fully determinate state. Okon and Sebastián's concluding line that 'We would be satisfied if our model turns out to be no worse than standard collapse models' offers little encouragement in this regard.

An alternative approach to the place of consciousness in the world is that of panpsychism, which pops its head above the parapet at various points in this volume. However, in their contribution, 'Quantum Mentality: Panpsychism and Panintentionalism', Acacio de Barros and Montemayor argue that the kind of mentality that is typically in play when it comes to the quantum context is not the phenomenalistic kind that panpsychism requires; rather, they claim, the intrinsic properties at the micro-level should be taken to be purely intentional. A key step in the argument draws on a form of perspectivalism, whereby such properties should be seen as tied to the mind of an observer. The latter not only appears on stage via the 'consciousness causes collapse' account but also, they suggest, is required by the inherent contextuality according to which the values of a given property are determined by the relevant context, where this must then be selected by the observer. In neither case, Acacio de Barros and Montemayor maintain, is phenomenal, rather than intentional, consciousness required (p. 96). That's as may be, but a sceptic may wonder why any form of consciousness should be required in either case! Leaving aside the concerns about the nature of causation when it comes to 'consciousness causes collapse', which Putnam and Shimony drew our attention to, when it comes to contextuality, the likes of Bohr ([1928]) were perfectly happy to acknowledge such context selection (which underpinned his response to the EPR thought experiment, of course), but they still insisted on the requirement of a 'detached observer'. Any role for consciousness in such an approach is severely attenuated.

Spontaneous collapse models return in the next chapter, 'Perception Constraints on Mass-Dependent Spontaneous Localization', in which Kent subjects to detailed analysis recent work on collapse times in the context of human perception. His conclusion is one that might apply to many of the contributions in this volume, namely, that the assumptions and approximations deployed are sufficiently questionable as to cast doubts on the result. In this specific case, it remains open whether mass-dependent continuous spontaneous localization models are viable.

The second section of the book kicks off with Goff's essay, 'Quantum Mechanics and the Consciousness Constraint', which shifts the interpretational focus to wave-function monism and asks whether this can satisfy the constraint mentioned in the title: any adequate theory of reality must entail that at least some phenomenal concepts correspond to reality. Following Wallace, Goff adopts a broadly structuralist approach in which everyday objects emerge as patterns in the global wave function. Facts about such objects are then taken to be analysable in terms of '*patterns of penetration resistance among regions of space*' (p. 123). However, he then avers, 'The trouble is that a description of the wave function does not logically entail that anything has the causal property of *resisting penetration*' (p. 124).

One option would be to agree with Ney ([2015]) that our understanding of such facts should be revised, but Goff argues that this runs afoul of the above constraint. However, the wave-function monist could simply refuse to meet this demand. So, the usual explanation for these 'patterns of penetration resistance'—as

manifested in the solidity of tables, for example—involves appeal to the Pauli exclusion principle, which is in turn underpinned by the anti-symmetrization of the relevant wave function. Such symmetry properties afford the wave-function monist all the resources she needs in this regard (see French [2013]).

Lewis, on the other hand, questions the very notion of 'experience' in play in these discussions, in his bluntly titled 'Against "Experience"'. Taking up Bell's ([1990]) famous dismissal of the term 'measurement' as a primitive in the standard formulation of quantum mechanics, Lewis argues that this should be extended to experience and its cognates, such as perception, observation, and indeed consciousness itself. Of course, as he then notes, a blanket condemnation of the use of such terms is hard to justify as there are some good examples in the literature, such as the afore-mentioned discussions of spontaneous collapse models in a neurophysiological context, where 'experience' is not treated as an unanalysable primitive. Bad cases occur when experience is dragged into the interpretation itself.

More interestingly, and rounding off the movie reference, there are the 'ugly' examples. One such is due to Hameroff and Penrose, whose contributions appear elsewhere in this collection, and who appeal to general relativistic considerations in the context of, yet again, a spontaneous collapse account. Here, however, they offer a specific hypothesis about how conscious experience arises, as we shall see. Lewis doesn't take this as undermining his core claim, since this hypothesis is quite separate from their interpretation of quantum mechanics itself (p. 150). What is important, he maintains, is to distinguish these different kinds of cases because the bad ones create confusion, not least by positing what amounts to an additional variable that in fact does no genuine explanatory work.

Ismael, however, takes a contrasting view in 'Why Physics Should Care about the Mind, and How to Think about It Without Worrying about the Mind–Body Problem'. Indeed, she insists that 'Because the evidence for our theories comes from experience [...] eventually we have to be able to bring the mind firmly under the scope of our physical theories and understand how human experience fits into the picture' (pp. 156–57). How this is actually to be done is not really tackled. As for the mind–body problem, in its 'hard' form, this can be ignored for the simple reason that 'if consciousness enters the problem space of physics, it does so by making a difference to the behavior of physical objects' (p. 167).

A more interesting analysis is offered by Skokowski in 'The Nature of Belief in No-Collapse Everett Interpretations'. Taking another interpretation, namely, Everett's ([1957]) 'bare theory' of quantum mechanics in which there is no collapse, Skokowski considers what happens when this bumps up against our understanding of the complex nature of belief states. Determinate observation utterances may be accounted for on this view by asking the observer a disjunctive question—'Did you get a determinate result for your measurement, either spin-up or spin-down (say)?'—that requires the observer to introspect their perceptual beliefs in order to evaluate the disjunction. However, Skokowski points out that answers to questions are formulated in a part of the brain that has to do with linguistic outputs and not introspection, and the former are not intentional. The latter would give us what is needed, except for the fact that with no collapse, there is no single belief state of the observer that can be extracted from the superposition with the singular content of being either spin-up or spin-down; that is, the issue of the superposition of consciousness must again be faced (p. 196).

Questions regarding the observer's perception also lie at the heart of the 'Wigner's friend' thought experiment, in which said friend is invited to open the box containing Schrödinger's unfortunate cat and

then asked what it is that they observe. Wigner's ([1962]) conclusion was that in order for a determinate answer to be given, collapse has to have occurred, and this could only be effected through the intervention of the observer's consciousness. In their thoughtful contribution, 'The Completeness of Quantum Mechanics and the Determinateness and Consistency of Intersubjective Experience', Silberstein and Stuckey consider a recent revival of Wigner's friend from which it has been concluded that we must give up on quantum mechanical completeness, locality, realism, or our intuitions about free will. Rejecting Everett's 'just take the theory as it is' approach, they adopt the 'principle-based' line that underpins the QBist and quantum information theoretic approaches. The idea here is to seek certain principles, akin to those we find in special relativity, from which the other features of quantum mechanics can be obtained (p. 200). In particular, the wave function is regarded not as representational but as epistemic and, more contentiously, a form of neutral monism is adopted, whereby physical entities are understood as 'contextually given manifestations' of a Jamesian 'unqualified actuality' (p. 248). Not surprisingly, perhaps, such a dramatic shift in framework undercuts the assumptions underlying the revised Wigner set-up.

Silberstein and Stuckey also extend their account to a 'toy' experiment that would test Bell's inequalities, with human subjects used to change the apparatus settings at the two ends of the set-up. As they note, Hardy in his contribution to the volume, 'Proposal to Use Humans to Switch Settings in a Bell Experiment', explores a realistic version of this very experiment. There, he argues that if there were agreement with the relevant Bell inequality, then the violation of quantum mechanics that this would involve would demonstrate that consciousness does play a special role in the world, one that Hardy suggests should be understood in dualistic terms. However, even he admits that, optimistically, the probability of such agreement is only around 1–2% (p. 310).

Silberstein and Stuckey, of course, reject dualism. Indeed, with their insistence that 'only subject–object is the basic unit of experience' (p. 251) and that physics is about the world of that experience, not some inaccessible, noumenal realm, their world-view sails close to that of Husserlian phenomenology. Perhaps they felt that to head in that direction would take them too far from the 'analytic' stance that pervades the entire volume. The exception is Bitbol's piece, 'The Roles Ascribed to Consciousness in Quantum Physics: A Revelator of Dualist (or Quasi-Dualist) Prejudice'. Here, he too rejects dualism but instead espouses the first-person standpoint of phenomenology. From this perspective, consciousness is a precondition of existence (p. 262), but not in a simplistically idealistic manner: physical objects have a certain 'immanent transcendence' insofar as there is more to them than we immediately perceive and they are 'given' to us in a way that allows for surprises (p. 263). This, then, 'invites us to see the existence of physical objects as an open problem rather than an uncontrovertible fact' (pp. 263–64) and thus can accommodate the problematic nature of the notion of 'physical system' within quantum physics. The issue now is how that accommodation might be developed.

Although Bitbol detects a phenomenological flavour in the Everettian interpretation, he argues that QBism offers the most consistent phenomenological approach, principally because it too insists on a first-person perspective (p. 274). Here 'state' vectors are regarded as nothing but Bayesian probabilistic valuations expressing the agent's subjective guesses. This is not to descend into idealism, however, and Bitbol nicely relates the 'participatory realism' of Fuchs and his co-workers with Merleau-Ponty's ([1968]) phenomenological account of embodiment (Fuchs and Stacey [unpublished]). However, hitching one's philosophical cart to a theoretical horse such as QBism incurs significant costs. In particular, certain features of quantum mechanics such as entanglement must be understood as derivative, not fundamental.

Those who are phenomenologically inclined but have less revisionary theoretical tastes might prefer some form of alternative partnership (see Berghofer and Wiltsche [2023]).

The third section of the volume presents a series of proposals for understanding consciousness via quantum mechanics, beginning with Penrose's 'New Physics for the Orch-OR Consciousness Proposal'. This is a long and wide-ranging piece spanning the author's well-known early work on Gödel's theorem, non-periodic tiling, and his defence of the gravity-induced objective reduction of the wave function. The 'orchestration' of objective reduction (OR) events then yields both classicality and consciousness, but since both gradual and instantaneous collapse models are deemed to be problematic, the collapse must take place retroactively. As Penrose acknowledges, this yields a 'strange' picture in which 'the "objective reality" of the situation is that the surviving member of the **OR** process was retroactively pre-determined by the "choice" that would *later* be made by the **OR** occurrence!' (p. 342).

The sense of 'objective' here is unclear, even more so when it is explained that it is through this retroactive collapse that determinate mental states arise. This occurs within the brain's (in)famous microtubules, certain of which possess symmetry features that may shield the large-scale quantum states required for macroscopic conscious experience. Penrose admits that much of this is highly speculative and grossly lacking in the necessary detail (p. 354), but in a companion piece, 'Orch OR and the Quantum Biology of Consciousness', Hameroff argues that the usual dismissive response that the brain is 'too wet and warm and noisy' to sustain such states fails to appreciate the highly heterogeneous nature of biological microenvironments. Indeed, he provides details of an 'underground' of 'quantum-friendly' structures, the geometries of which are then claimed to support a 'decoherence-free subspace' (p. 378) that could be conducive to the Orch-OR mechanism. Responding to recent criticisms, Hameroff maintains that the model at least has the virtue of falsifiability, while remaining unfalsified so far. However, it is worth noting that the details have shifted: from the suggestion that some form of Bose–Einstein condensation was involved to an alternative mechanism, also criticized, and so on. Perhaps the approach is better thought of in terms of a Lakatosian research programme, so the question then is whether it should be regarded as progressive or degenerating. Certainly, the 'hard core' comes with high 'buy in' costs.

One might adopt a similar view of the Bohm–Hiley account, presented here by Hiley himself, together with his collaborator Pylkkänen, in 'Can Quantum Mechanics Solve the Hard Problem of Consciousness?'. This is another long piece that also presses for a radical change in our understanding of reality by allowing 'non-mechanical, organic and holistic concepts such as active information to play a fundamental role' (p. 413). Again, the details will be familiar to many: the quantum potential of the Bohm interpretation is understood as 'enfolding' information about the environment (such as the precise nature of the slits in the two-slit experiment) so as to organize the behaviour of the associated particle. How this organizing power might actually work isn't clear, despite the variety of analogies deployed.

The argument then skips from that enfolding of information by the quantum potential (p. 426) to the association of meaning with the former, to the conclusion that meaning must be involved in particle behaviour. This is a two-way street: configurations of matter possess meaning and meaning can organize matter, especially at the quantum level. What we end up with is a form of panprotopsychism, according to which quantum particles have certain primitive mind-like qualities, albeit falling short of consciousness (p. 433). Proceeding in the other direction, active information at the quantum level also has protophenomenal properties so that when quantum systems are arranged in the right kind of structure, as in the brain,

phenomenal properties emerge. Indeed, Hiley and Pylkkänen take us even further, suggesting that the holism inherent in quantum theory generates a kind of 'cosmopsychism'! Some may feel that even after having paid the high entrance fee, they'll want to get off the ride before reaching this point.

This account also appears in Seager's 'Strange Trails: Science to Metaphysics' as an example of a response to the appeal for greater 'metaphysical intelligibility' in our world view, something that has been lost since Leibniz: 'Like Leibniz, Bohm moves towards a view which puts mentality, as the bearer of intrinsic information, as a fundamental feature of the world, whose attributes go at least some way towards providing the metaphysical intelligibility needed to buttress and complete the mathematical intelligibility so evident in modern physics' (p. 478). However, recent work in the metaphysics of science suggests that we have more options than a return to the eighteenth-century aphorisms of the *Monadology* to secure the resources we need for greater 'intelligibility'.

Leibniz's principle of identity of indiscernibles also features as a core principle of Smolin's relational hidden variables theory, which attempts to unify quantum mechanics and general relativity (although the principle is of course contentious in the quantum domain). His chapter, 'On the Place of Qualia in a Relational Universe', further expands on this programme, with the idea of the universe as 'constructed from nothing but a collection of *views* of events' (p. 483). Moments of consciousness are then taken to be aspects of certain of those views, namely, those that are unprecedented and/or unique. These are physically distinguished from their common-or-garden counterparts 'most likely' because of their greater complexity, which precludes them from being copied.

What this gives us is a 'restricted panpsychism' that explains why qualia are never perceived singly but only bundled together with others, since each conscious perception correlates with a 'view'. However, this is admitted to being only a 'hypothesis' and much of the presentation here is 'hand-wavey' at best. The same can be said of the concluding paragraph on making contact with the relevant neurobiology, where now it is the 'bilayers of phospholipid molecules' forming the neuron's membrane that are proposed as containing the relevant quantum effects.

By this point one can't help but feel sympathetic to Eddington's famously proscriptive stance in the 1920s when he suggested that a sign saying 'Work in Progress: Keep out' should be nailed above the door to physics departments! But it's not just more experimental work that is needed here—the underdetermination arising from the choice of such different interpretational frameworks with regard to both quantum theory and consciousness may also indicate that a radically alternative approach is needed. Nevertheless, this rather inhomogeneous but suggestive collection does at least bring the issues involved back onto centre stage.

Steven French
University of Leeds
s.r.d.french@leeds.ac.uk

References

Berghofer, P. and Wiltsche, H. [2023]: *Phenomeonology and QBism: New Approaches to Quantum Mechanics*, New York: Routledge.

French, S. [2013]: 'Whither Wave-Function Realism?', in D. Albert and A. Ney (*eds*), *The Wave Function*, Oxford: Oxford University Press, pp. 76–90.

Bell, J. S. [1990]: 'Against "Measurement"', *Physics World*, **3**, pp. 33–40.

Bohr, N. [1928]: 'The Quantum Postulate and the Recent Development of Atomic Theory', *Nature*, **121**, pp. 580–90.

Everett, H. [1957]: '"Relative State" Formulation of Quantum Mechanics', *Reviews of Modern Physics*, **29**, pp. 454–62.

Fuchs, C. A. and Stacey, B. [unpublished]: 'QBians Do Not Exist'.

Maudlin, T. [1995]: 'Three Measurement Problems', *Topoi*, **14**, pp. 7–15.

Merleau-Ponty, M. [1968]: *The Visible and the Invisible*, Evanston, IL: Northwestern University Press.

Ney, A. [2015]: 'Fundamental Physical Ontologies and the Constraint of Empirical Coherence: A Defense of Wave-Function Realism', *Synthese*, **192**, pp. 3105–24.

Putnam, H. [1961]: 'Comments on the Paper of David Sharp', *Philosophy of Science*, **28**, pp. 234–37.

Shimony, A. [1963]: 'Role of Observer in Quantum Theory', *American Journal of Physics*, **31**, pp. 755–77.

Wigner, E. [1962]: 'Remarks on the Mind–Body Question', in I. J. Good (*ed.*), *The Scientist Speculates*, Portsmouth, NH: Heinemann, pp. 284–302.

Cite as

French, S. [2023]: 'Shan Gao's *Consciousness and Quantum Mechanics*', **2023**