Consciousness and no-collapse physics

Christopher J S Clarke
School of Mathematics, University of Southampton
Southampton, SO17 1BJ
United Kingdom
cclarke@scispirit.com

Abstract

The study of consciousness was given its modern direction by David Chalmers' introduction of the "hard problem" concept in 1995. It has since become a crucial bridge between psychology and physics. In this period the problem has become more tractable through work in cognitive psychology by Philip Barnard and progress in the foundations of quantum theory. Both of these are drawn on in this paper, which gives an overview of the applicability of quantum theory to the "hard problem". It presents a brief outline of the two-way interaction between quantum processes and consciousness, emphasising the physical correlates of consciousness but taking into account the complementary methodology required for the investigation of consciousness itself. A framework is presented for defining the role of consciousness both in psychology and in the foundations of quantum theory.

Keywords

Quantum theory, Consciousness

1. Usage: the scope of "consciousness"

This word is used in many different ways. Following Velmans[1], I am taking "consciousness" to mean the active facility for experiencing in an individual system. It is in this sense that Chalmers[2] defined the "hard problem of consciousness" as "How can we explain why there is something it is like to entertain[such things as] a mental image, or to experience an emotion?" This sense of the word does not refer to information processing (which could equally well happen "in the dark" as Velmans puts it) nor to why we are conscious of some things and not others, and it is not restricted to the reflexive consciousness of one's self. It is the bare fact of awareness . In this paper it does not refer to a universal consciousness.

This definition will be relevant not only to human beings, but also to other systems, including non-human organisms. Although, as argued by Nagel [3], it may be impossible to *know* "what it is like to be" a non-human entity, so that we may only be able actually to verify the presence of consciousness in human beings (through empathy, for example), we can still apply our understanding of consciousness to a wider range of systems . The point is not, whether we can in practice know this directly, but whether or not there can theoretically exist, for some particular system, a "what it is like to be" in the sense of bare awareness as defined above.

Various different sub-systems of organisms can also, in principle, be conscious and so for complex organisms like humans, we need to distinguish the subsystem on which the human can verbally report from those that are inaccessible to report. The former is the "seat" of "my" consciousness in

traditional terminology (for Descartes, the pineal gland). Consideration of multiple consciousnesses[4] could be relevant to considerations of systems of organelles in single cells.

2. Mind and the world

To avoid misunderstanding I should next clarify my philosophical position. I accept Kant's core proposition as interpeted by Savile[5] that the world *as it is known by us* is structured *a priori* by our capacities of knowing. I would also, unlike Kant, recognise that this capacity changes with time through changes in human culture and genetics.

Several theories imply that we have two different primary modes of knowing, imparting a dynamical polarity to the human mind, which, following Kant, implies that the world manifests to us two different aspects. The polarity theories for which there is strong evidence are the "interacting cognitive subsystems" (ICS) model of Teasdale and Barnard[6] and the recent cerebral hemisphere model of McGilchrist[7]. The former uses data from experimental psychology to support a scheme that includes and refines several psychotherapeutic models, while the latter is based on a detailed survey of behavioural and neurological evidence that the two hemispheres of the brain induce a polarity of knowing strikingly similar to the ICS model. Each sheds light on the other.

According to Teasdale and Barnard[6] we are governed, at the top level of our mental organisation, by two distinct meaning-making "interacting cognitive subsystems". One (the "implicational" subsystem) is concerned with the significance for the self of its overall context, drawing immediately of our sensations. It deals with what *concerns* us, including monitoring threats and opportunities, and with relationships, in the sense of our meaningful connections both with other beings and within ourselves. The other subsystem (the "propositional") is concerned with analysing experiential data from the implicational subsystem – abstracting from it general concepts (including the self-image) and linking them into conceptual propositions, which are in turn fed back to the implicational subsystem, thus enabling thinking to be reflective. It is closely connected with speech but has no direct contact with the senses. In terms of logic, the propositional is the basis of rational thought while the implicational is pre-rational. Each subsystem has its own memory store with characteristic access modes and transfer times between them.

McGilchrist's model[l7] is similar, associating the higher level cognitive functions of the left and right hemispheres with properties very close to those of the propositional and implicational subsystems, respectively. He stresses that, in their cognitive functions, the right hemisphere (implicational) deals in whole entities, the left in abstractions.

I have in the previous section defined consciousness by distinguishing it from its content and its physical correlates, leaving only the ground of our awareness. Consciousness in this sense delivers the immediate "presence" of the world as it appears to us (in the language of the debate between Steiner[8] and Derrida) rather than its conceptual structure. In terms of the polarity described above, consciousness is therefore associated strictly with the implicational subsystem, so that we cannot grasp consciousness by rationality alone: only the implicational subsystem can deliver the "presencing" which constitutes awareness. This imposes a methodological requirement on the study of consciousness, which essentially requires a non-verbal, non-analytical investigation to complement a standard scientific investigation of the physical correlates of consciousness. There is,

of course, a very extensive literature on the former (particularly within Buddhism) based on detailed personal , internal non-verbal investigation. It requires some personal acquaintance with actual practice for its comprehension; given this, it is making a large contribution to this area.

The main problem now is that of relating what we know about consciousness from internal investigation to the physics of the seat of consciousness. Although consciousness itself is not governed by logic in the formal sense, some logics are suited to relating consciousness to physics. The Chilean analytic psychologist Ignacio Matter Blanco[9], encountering the same problem when trying to describe how the *un*conscious appeared to be operating, noted that it was as if it obeyed an alternative logic, which he termed bilogic. Some of its features are highly relevant to the issue here: it is a context-dependent logic (a feature already implicitly used in quantum theory in the form of topos logic[10]) and it does not have a well-behaved operation of negation. I shall return to these points in the final section.

3. Quantum and collapse

The following are the main reasons for looking to a quantum description in order to provide the physical correlate of consciousness.

Q1. Consciousness seems to draw percepts into a unity (the perceived world) of co-extensive qualities. On the physical side, the "quantum state" (which is linked with space but not extended over it) seems a more appropriate reflection of this than does the "atomistic" world view of classical physics.
Q2. Quantum physics provides a general language ("proposition", "state", "observation" …) that allows a closer mirroring of the perspective of consciousness.
Q3. As an added bonus, quantum physics has problems of its own that could be resolved by linking it with consciousness.

As there are confusingly many forms of quantum physics in use, I will outline what I consider its main features.

The theory owes its basic structure and ideas to Niels Bohr, who rooted quantum physics in laboratory practice, stressing the need for connecting quantum physics with the basic quantitative concepts on which classical physics was founded (position, mass and momentum in particular) and the relations between them. The later formalisation of this by von Neumann still forms the basis of the main-stream approach. In this, the defining core of quantum physics is a process of *observation* (or *measurement*) having three stages: the preparation, or selection, of a system in a well defined state; the placing of the system in an observing apparatus which defines a particular context (what it is that is going to be measured); and the result of the measurement, for which the quantum formalism supplies the probabilities of obtaining all possible values. These stages are illustrated below, with some of the alternative terminology (corresponding to alternative conceptualisations) added.

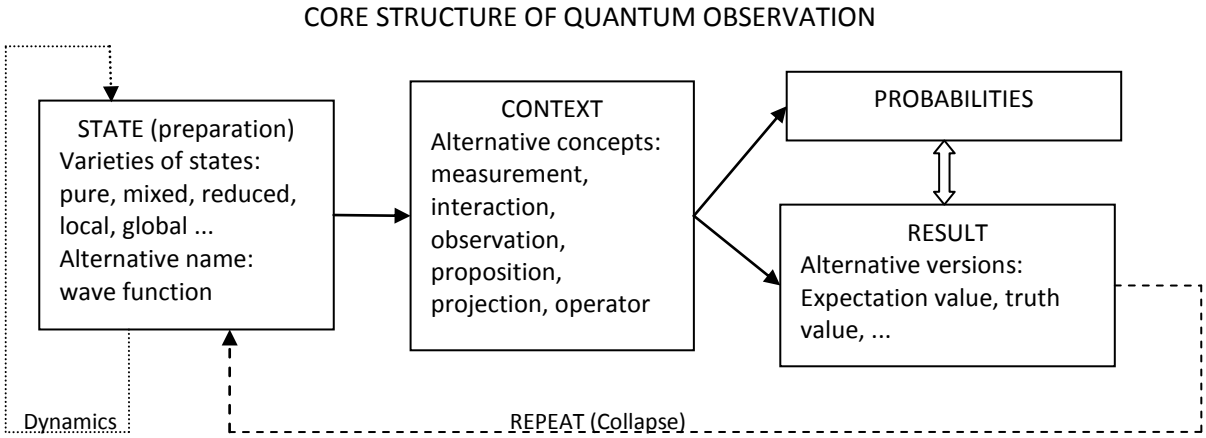

Figure 1: alternative terminologies referring to quantum observation

If the system is then fed into a repetition of this process, the probabilities will be different, indicating that the state has been changed by the observation. This is quantum *collapse.* Alternatively, the system may be left alone to evolve through a *dynamic* which exactly mirrors the dynamics of Newtonian theory.

Since most of the cosmos does not contain physics laboratories, the process just outlined has to be generalised for application to other fields, which has given rise to many different approaches. The nature of "collapse" is particularly controversial, with four main strands of thought (omitting the "many worlds" interpretation which just replaces "collapse" by "splitting" into universes, leaving all the problems unchanged[1112] and Bohm's theory[13] which is viable but too different to consider here):

C1. Minimal no-collapse quantum theory[14]. There is no such thing as collapse. In a measuring apparatus the system is coupled to a detector which is physically large and so therefore interacts with the large-scale environment. This interaction feeds back to the system and modifies it in a random way because of the coupling. As a result the *reduced state* of the system (the mixed state that expresses probabilities for a subsequent measurement on the system alone) is an ordinary statistical distribution, as might be used when rolling a dice. Performing a measurement then discloses which actual state is present out of this distribution.

An objection to this is that the conclusion is achieved by a slide of meaning from quantum state to statistical state and back to quantum state, which hides the fact that the basic problem has not actually been solved, but only spread out over the universe.

C2. There is an additional physical process involved that performs the collapse. An early proposal by Ghirardi, Rimini and Weber has problems[15] though some are being successfully tackled[16]. The version relevant to consciousness is that proposed by Penrose based on quantum gravity which I discuss in the next section.

C3. There is no such thing as collapse: the quantum state of the universe, with everything including laboratories within it, evolves by a deterministic dynamics. This quantum universe then determines the probabilities for either (a) all possible sequences of experiences by "brains", or collections of "brains", in the universe; or (b) all possible instances of a sensation by a "brain" in the universe.

Case (a) allows one to derive from this the probabilities for individual experiences by considering sequences of repetitions of experience, using an extension[17] of the statistical argument used by Everett for the many worlds model. The problem with this is that it is very unclear how one generalises the particular case of a human brain so as to formalise the essence of the sort of thing that can have an experience. I take this up in the next section.

C4. There is a non-physical process, such as awareness by a "soul" or by "consciousness" as a non-physical substance, that produces the collapse.
The problem here is that this cannot be considered until we have a coherent metaphysics of the soul, including how it can interact with matter without having a physical aspect. In the absence, in my opinion, of such a metaphysics (despite many valiant attempts), this is an "and then a miracle happens" argument.

Out of these, C1 is well established. The basic mechanism involved is now well understood, and is clearly important in transforming the non-standard logic of quantum physics into the standard logic of classical mechanics with a statistical distribution of outcomes. What it does not do, however, is to explain how and why one particular outcome from a superposition of possibilities actually emerges. Something else is required in addition if one is to go beyond a purely laboratory setting.

4. Solving the quantum and consciousness problems

4.1. Penrose's theory

I will now put the preceding sections together to explore whether, by coupling quantum physics with consciousness (in the sense in which I have defined it), the problems in both these areas can be resolved.

The leading candidate in the field at present is the theory of Hameroff and Penrose[18], which uses option C2 of the list in the preceding section. I will briefly enumerate here the reasons why I am not fully convinced that this solves the consciousness problem, before going on to describe an alternative.

1. The mechanism they describe is claimed to produce a non-algorithmic information processing system in the brain. If correct, this would still not supply any connection with our immediate awareness, which is a matter of participating in experience, not processing information. Penrose's mechanism may explain why life is so clever, but not the hard problem as I define it here.
2. In any case, there does not seem to be any reason why the collapse described by Penrose should involve information processing at all, rather than random selection. I would agree with him that humans are capable of non-algorithmic information processing, but how does collapse actually achieve a non-algorithmic processing of information (bearing in mind that quantum computers are algorithmic)?
3. The argument that gravitation produces collapse[19] needs a lot more investigation than it has so far received. Penrose makes an analogy between, on the one hand, the mathematics of gravitating superpositions which fail to have a stationary state and, on the other hand, the decay of an unstable nucleus: but the latter *presupposes* the occurrence of a collapse, so we cannot use this as an independent argument for collapse. The core of Penrose's argument is

that when space-time loses its symmetry it is not possible (in the semi-classical model being envisaged[19]) to superpose the states involved; but the failure of this model would seem to argue for the introduction of a fuller theory, not for the appearance of a completely new collapse process intervening in quantum physics.

Here I propose using the principle C3(a) of the previous section as the main focus for understanding consciousness. Many of the arguments which follow are equally applicable to the Penrose-Hameroff model: these are not entirely exclusive alternatives.

4.2 How consciousness can replace collapse

I will start with the ideas of Don Page[20], who postulates that the quantum state of the *universe* (as a completely homogeneous model) undergoes a dynamical evolution with no collapse. At any moment of time (time is well defined in this model), the state can be expanded in any one of a vast numbers of ways as an orthogonal superposition of other states. If one of the components of such a superposition has the properties which correspond to its being the state of a conscious entity , then the experience of this entity is a possible experience of a universe. So the properties that define consciousness determine the possibilities for what the universe can look like to a conscious being. This explains, he suggests, the *world* (as distinct from the universe) as we see it as one possible conscious snapshot of the universe, that is selected out of the vastness of all possible components by criteria for consciousness.

Page applies this argument to just one moment of sensation by one conscious entity. If the conscious entity concerned has, in addition to consciousness, the sort of content that we call "memories", then this moment of awareness will seem to it as if it is part of a historical sequence of events. This is a rather extreme form of solipsism, where not only might I be the only conscious being in the universe, but my life might consist only of this one instant! I consider, however, that if consciousness is presencing, as noted above, then it requires other beings to be co-present and myself to have been present in the past. I will therefore modify Page's rather austere proposal to take this into account, by regarding the perceived world as the world of a community of conscious entities which are enduring in the sense of having repeated moments of awareness.

At the formal quantum level, this modification has been developed by Hartle[21], following his joint work with Murray Gell-Mann in the early days of quantum cosmology, based on the "history interpretation" of quantum theory. His formalism enables one to associate a probability with any such enduring community in which each moment of awareness contributes a factor to the total probability. The factor depends on the relationship between the moment's own content and that of its predecessors. If, for example, my awareness a few seconds ago (time t_1) had included my sitting in my house, but in my awareness now (time t_2) the house has disappeared , this would assign a very low probability factor to the overall tally.

The mathematical process for finding the probability p of a collections of moments of awareness is based on the standard decoherence functional of quantum history theory. It involves[22]:

- associating with each moment of awareness (including the qualitative "what the world should be" raised in section 4.4 below) a projection P on the state space of the seat of consciousness and extending this to a projection q_i at the relevant time t_i on the whole universe;

- using the dynamics of the universe to map all these projections back to an arbitrary moment of time t_0 of the universe (in order to compare them on the same footing);
- multiplying them together in their correct chronological order to obtain an operator Q;
- and finally applying Q to the state ρ_0 of the universe at time t_0 using the usual formula (p = tr[$Q\,\rho_0 Q^{\dagger}$])for the probability p of a sequence of quantum measurements.

This gives the same final result as the "consciousness collapses the universe" position of C4 in the previous section but, crucially, without the collapse. What happens in this proposal is that consciousness *selects* from the universe a range of possible worlds and assigns a probability to each. The quantities p should strictly be called quasi-probabilities because they only satisfy the axioms for probability *when decoherence is taking place.* Selection generates the world, while decoherence makes it classical.

4.3 What things are conscious?

The question now is, what are the criteria for consciousness? I have already severely limited the restrictions these criteria can impose by defining consciousness as a *sui generis* quality distinct from the objective *content* of an entity's consciousness (which might include information processing) and also by claiming that we can only recognise consciousness through the non-verbal means of the implicational subsystem. Thus it is hard, perhaps positively misguided, to draw a line between those entities that are conscious and those that are not. Several writers have recently considered this dilemma[23,24,25], and come to a panpsychist position: that everything is conscious. The meaning here is ambiguous, however, varying between the idea that there is a diffused consciousness throughout the universe (a concept distinct from my usage here which refers to individual consciousness) and the idea, relevant here, that every "thing" is conscious.

This brings us to the question "what is a thing?", the title of a lecture series by Heidegger[26] and, echoing this, a paper[10] by Döring and Isham on an alternative that bypasses quantum collapse. Tangentially to both of these, my question here is , given the holistic way in which quantum states and the quantum field encompass the whole of space or space-time, what structures are there within the basis of modern quantum physics (viz. quantum field theory) that correspond to (comparatively) discrete individual entities, and which include living organisms and organelles? I suggest that the answer lies in *coherence* as the word is used by Mae-Wan Ho[27]: a region that holds together dynamically, whose parts respond in harmony with each other.

From a physics point of view there are, however, two distinct senses of "coherence". One is the opposite of "decoherence" (see C1 in section 3): a decohered region has separate states in separated parts, so that the total state is just the tensor product of the states in its parts. Coherence as the opposite to this occurs when the states of all separated parts are entangled. The other sense of "coherent" occurs in optics, including quantum optics, where the phases of the light in two separate regions are in step. But this can happen coincidentally without any interaction between the two parts – as when, for instance, beams of light from two different lasers set up at exactly the same frequency are brought together and produce an interference pattern[29.] My proposal for a "thing", therefore, is a local quantum state whose spatially separated parts are entangled. While the concept is quite intuitive, the actual definition of the degree of coherent entanglement is rather complex[22:A2.3]. Importantly, this definition covers exactly the situation required for microtubules in

the approach of Hameroff and Penrose; so we are considering the same sort of entity, but from somewhat different viewpoints.

4.4 What does consciousness do?

Though it might turn out that consciousness, as defined here, is an epiphenomenon, our subjective experience is that consciousness does something: we not only enjoy our world , but we are agents in shaping it. Our action through consciousness, however, is unlikely to be an overriding of the dynamics of quantum theory, or even a selection of one outcome of a quantum measurement that otherwise would be probabilistic, because this would manifest itself clearly as the parapsychological phenomenon of "psychokinesis", known experimentally to be at best an extremely weak effect. From the implicational point of view (and noting the results of Batthyany[28] on extending Libet's work) consciousness can bring to the world not only a subjective "what the world is like", but also a qualitative "what the world should be". This enables consciousness to choose (or at least weight the choice of) the proposition *P* in Hartle's formalism, though not the probability of the outcome. I shall call this *asserting P.* In this case the non-occurrence of *P* does not imply the occurrence of *not-P,* in keeping with the non-classical logic of the implicational noted in section 2. This then raises the following possibilities for the action of consciousness[30].

A1. The seat of consciousness can maintain its own coherence by repeatedly asserting its own internal entanglement, employing the well established quantum Zeno effect. (This implements Spinoza's "connatus" stressed by Mathews[23] as an ingredient of pan-psychism.) This in turn makes possible a larger seat of consciousness which can act physically as a quantum computer.

A2. In addition, where the seats of consciousness of two people become momentarily entangled, the same mechanism can maintain the entanglement, establishing a correlation between the actions of the people (though without information transfer)[30].

A3. Where quantum computation is taking place, consciousness can assert projections on superpositions of classically incompatible large-scale situations, enabling creative insights impossible in conventional logical processes.

A4. We can note that Hartle's process is symmetric in time, implying that future assertions can affect the present. This produces apparently "parapsychological" effects[30].

5. Conclusions

I have presented a theoretical framework based on ICS theory, Hartle's interpretation of quantum theory, a panpsychist approach to consciousness and a coherence criterion for the seat of consciousness. There is scope for variation in all these components in the light of future research. Items A1, A2 and A4 above suggest immediate opportunities for further theoretical and experimental research. The most radical proposal for future work is the suggestion that a full solution to the "hard problem" requires the coupling of the two different methodologies of interior investigation and mainstream science, underpinned by cognitive psychology.

References

1. Velmans M. Understanding Consciousness. London: Routledge; 2000.

2. Chalmers DJ. Facing Up to the Problem of Consciousness. Journal of Consciousness Studies. 1995;2:200-219.
3. Nagel T. What Is it Like to Be a Bat? Philosophical Review. 1974;83(4):435-50.
4. Lockwood, M. Mind, Brain and the Quantum: the compound 'I'. London: Blackwell; 1990.
5. Savile A. Kant's Critique of Pure Reason: An Orientation to the Central Theme. Malden, MA: Blackwell; 2005.
6. Teasdale JD. Barnard PJ. Affect, Cognition and Change: remodelling depressive thought. Hove: Lawrence Erlbaum Associates; 1993.
7. McGilchrist I. The Master and his Emissary: the divided brain and the making of the Western world. New Haven and London: Yale University Press; 2009.
8. Steiner G. Real Presences. London: Faber and Faber; 1989.
9. Matte Blanco I. The Unconscious as Infinite Sets. London: Karnac; 1998.
10. Döring A, Isham CJ. What is a thing? Topos theory in the foundations of physics. arXiv:0803.0417 (http://arxiv.org/abs/0803.0417); 2008. To appear in "New Structures in Physics" editor R. Coecke.
11. Drescher GL. Good and Real: demystifying paradoxes from physics to ethics. Cambridge MA: The MIT Press; 2006.
12. Clarke CJS. Reviews: Good and real [document on the Internet]. Last updated 2010 June 23. Available from http://www.scispirit.com/drescher.htm.
13. Bohm D, Hiley BJ. The Undivided Universe: an ontological interpretation of quantum theory. London: Routledge; 1993.
14. Schlosshauer M. Experimental motivation and empirical consistency in minimal no-collapse quantum mechanics. Ann. Phys. 2006;321:112-149.
15. Blood C, Difficulties with collapse interpretations of quantum mechanics. NeuroQuantology. 2009;June 7(2):302-313.
16. Tumulka R. A relativistic version of the Ghirardi-Rimini-Weber model . Journal of Statistical Physics 2006;125:821-840.
17. Clarke CJS. Quantum theory and cosmology. Philosophy of Science. 1974;41:317-332.
18. Hameroff S, Penrose R. Conscious events as orchestrated space-time selections , Journal of Consciousness Studies. 1996; 3(1):36-53.
19. Penrose R. The Road to Reality: a complete guide to the laws of the universe. London: Jonathan Cape; 2004. (pp 846-856)
20. Page DN, Mindless sensationalism: a quantum framework for consciousness. In Smith Q, Jokic A, editors, Consciousness: new philosophical essays. Oxford: Oxford University Press; 2001.
21. Hartle J. The quantum mechanics of cosmology. In Coleman S, Hartle P, Piran T, Weinberg S, editors, Quantum cosmology and baby universes. Singapore: World Scientific; 1991.
22. Clarke CJS. The role of quantum physics in the theory of subjective consciousness. Mind and Matter, 2007;5(1):45-81.

23. Mathews F. For Love of Matter: a contemporary panpsychism. Albany: State University of New York Press; 2003.
24. Skrbina D. Panpsychism in the West, Cambridge MA: Bradford Books; 2005.
25. de Quincey C. Radical Nature: rediscovering the soul of matter, Montpelier VT : Invisible Cities Press; 2002.
26. Heidegger M. What Is a Thing? Translated by W. B. Barton, Jr. and Vera Deutsch. Chicago: Henry Regnery Company; 1967.
27. Ho M-W. The Rainbow and the Worm: the physics of organisms (2nd Edn). Singapore: World Scientific; 1998.
28. Batthyany A. Mental causation and free will after Libet and Soon: reclaiming conscious agency. In Batthyany, A, Elitzur A editors. Irreducibly Conscious: selected papers on consciousness, Heidelberg :Winter; 2009.
29. Magyar G, Mandel L. Interference fringes produced by superposition of two independent maser light beams. Nature 1963;198:255.
30. Clarke CJS. A new quantum theoretical framework for parapsychology. European Journal of Parapsychology 2008; 23(1):3–30.