Can quantum mechanics explain "free will"

Discussion in *Trends in Cognitive Sciences*.

My paper:
Funda-Mentality: Is the conscious mind subtly linked to a basic level of the universe?
Trends in Cognitive Sciences 2(4):119-127 (1998), was followed by an exchange focusing on the problem of fr will (and other issues) between me and two critics, Emmet Spier from the Centre for Computational Neuroscien and Robotics, University of Sussex, Falmer, Brighton, UK, and Adrian Thomas, Department of Zoology, University of Oxford, South Parks Road, Oxford, UK. They get the last word in this exchange, but I'll counter later.

Response from Emmet Spier and Adrian Thomas

Recently there have been a number of attempts from students of mathematical and physical backgrounds to re-establish a dualistic approach to mind and behaviour. Most notable has been the work of Penrose[1]; however, his basic idea has origins in the work of Popper and Eccles[2]. In the preceding article, Hameroff provides an exampl of the acceptance by this group of authors of the division that Descartes made between mind and body. Unsatisf (probably with good reason) with any of the mechanistic accounts of 'our mental experience' he makes a specifi proposal that a purely speculative scientific theory (quantum gravity) provides the conduit between mind and body. The essential premise from which an appeal to quantum gravity might make sense is Penrose's argument that the mind can perform non-computable operations. Certainly, quantum mechanics supposes that the world h the incomprehensible property of existing in a superposition of possible states and every so often plumps for on or another. If one follows Penrose and Hameroff and assumes that the hypothetical quantum gravity will also ha the same property, then such a phenomenon might offer a route for non-computability in the brain. We should a reflect upon the example they provide to establish that the mind performs non-computable operations. Penrose'claims that, although a human mind can comprehend the proof of Gödel's theorem, such an understand cannot come from within a formal computational system. This may well be a sensiible argument against some claims (although Sloman[3] provides a thorough defence). However, it hardly establishes non-computability as a yardstick with which to measure mental explanations. Indeed, just as our appreciation of the truth of Gödel's theorem comes from taking a wider, and less exact, perspective than just the formal proof, there seems to be (as Turing[4] thought) no *a priori* reason why a purely mechanistic system could not also bring knowledge from outs a formal system (general knowledge indeed) to aid its 'understanding'

However, Hameroff is also concerned with terms like 'free will' which any mechanistic system would find hard provide. Certainly something indeterminate, like quantum gravity, offers more hope for 'free will' Penrose and Hameroff[5] propose that the microtubules of the cell cytoskeleten could be used to bring some of the weirdness o quantum mechanics into the classical realm. Their idea is expressed thus: 'the picture I have is that for a while, these quantum computations go on and they keep themselves isolated from the rest of the material for long enou 'perhaps something of the order of nearly a second 'that the kinds of criteria I was talking about take over from t standard quantum procedures, the non-computational ingredients come in, and we get something essentially different from standard quantum theory.'SUP>1 Penrose and Hameroff's belief[6,6] is that systems of microtubule might sustain large-scale quantum-coherent activity, with 'individual OR occurrences constituting conscious events.' Because microtubules are tubes, Hameroff suggests that their insides could be in some way isolated fro the random fluctuations (heat) in the environment, which is, of course, crucial to maintaining quantum coherenc for extended periods of time. However, we do not find such arguments convincing, because microtubules are in

fact dynamic entities existing in a balance between polymerization and depolymerization[7]. Notably, individual tubulin dimmers are constantly being added and removed from the open ends of the microtubules and individua microtubules have a half-life of about 10 minutes.

To support his argument that microtubules might somehow be involved in the 'mind' rather than simply in neuro and the nervous system, Hameroff points to the complex behaviour of single-celled organisms. Unicellular organisms, by their nature, lack nervous systems, yet they are capable of complex behaviours, such as chemotax and object avoidance. Hameroff and Penrose both cite the ciliate paramecium, which has a cell cytoskeleton ric microtubules. These microtubules act as skeletal elements anchoring the cilia, and resist the forces generated by the cilia as they propel the animal through the water with powerful 'rowing'strokes. Hameroff suggests that the microtubules of a paramecium's cytoskeleton might also play a role in performing the computations necessary f its complex behaviour, by accessing the quantum mechanical realm in the same way as he proposes that the bra does. (Presumably because he supposes that paramecium compute by using the 'OR'procedure noted above, Penrose must also believe that paramecium have 'free will'and 'experience') However, through careful experimental work and detailed computational modelsg,[8,9], the computational processing systems underlying th complex behaviour of unicellular organisms are now rather well understood. Protein molecules act as the computational elements in living cells, performing computations by their complex chemical interactions. Even bacteria are capable of complex behaviour and, in the case of clonal bacteria, can achieve much higher degrees behavioural coordination than paramecium ever shows. The complex behaviour exhibited by bacteria (for example, chemotaxis) can be modeled from 'first principles'from the chemistry of the complex interacting molecules and the reactions that drive metabolism. As Bray[8] notes, 'In unicellular organisms, protein-based circ act in place of a nervous system to control behaviour.'Microtubules are nowhere to be seen and, as such, can ha be the necessary elements required to do whatever computing is needed for behaviour.

It is worth noting that, in Hameroff and Penrose's theory, nothing is specified to do the work of the mind. That achieved by an amorphous (immaterial?) quantum entity that interacts with 'Platonic logic embedded in spaceti The implications of such an opinion are unclear. One possible reading would imply that because there must, by definition, be only one Platonic logic, then all agents performing under it must hold the same opinions-just as al mathematicians agree on πor G'/FONT>del's theorem. Neither is it clear how the appeal to such a Platonic spac furnishes an explanation for 'free will'or 'experience'that is any more informative than a mechanistic attempt.

Clear though it is that consciousness is, for many, a mystery needing explanation, Hameroff and Penrose's atter to offer the microtubules within neurons the same place in their theory as Descartes'use of the pineal gland in hi is yet again making the brain a (now quantum) antenna for the mind. The new conduit is identified by drawing from new and speculative ideas in physics. This can be a dangerous form of logic 'the fact that two areas of scie are not understood does not imply that they are connected. Irrespective of how exciting these new ideas in phys are, it seems a little churlish of their proponents (as can be the wont of physicists) to try to explain all phenome from their level of explanation. We fear the necessary insights are some distance away.

References

1. Penrose, R. (1989) The Emperor's New Mind, Oxford University Press.
2. Popper, K. and Eccles, J. (1977) The Self and its Brain, Springer-Verlag.
3. Sloman, A. (1992) The Emperor's real mind Artif. Intell. 56, 355-396.
4. Turing, a.M. (1950) Computing machinery and intelligence Mind 49, No. 2236, 433-460.
5. Hameroff, S.R. (1998) 'Funda-mentality' is the conscious mind subtly linked to a basic level of the

universe? Trends Cognit. Sci. 2, 119-124.
6. Penrose, R. (1996) The Large, the Small and the Human Mind, Cambridge University Press. Alberts, B. etal. 91994) The Molecular Biology of the Cell, Garland Publishing. Bray, D. (1995) Protein molecules computational elements in living cells. Nature 376,307.
7. Bray, D. and Bourret, R.B. (1995) Computer analysis of the binding reactions leading to a transmembra receptor-linked multiprotein complex involved in bacterial chemotaxis. Mol. Biol. Cell 6, 1367-1380.

Reply to Spier and Thomas from Stuart Hameroff

I thank Drs. Spier and Thomas for their interest and forthright criticism. At first glance, a link between consciousness and a basic level of the universe is indeed a strange departure and their skepticism is understandable. However, the 'hard problem'of conscious experience demands new ideas. The Penrose-Hamero Orch OR model encompasses not only physics, but also philosophy, neuroscience, cognitive psychology, comp science, molecular biology and evolution.[1-3] Such an integrated, multilevel assault is required for any serious attempt. However, rather that dualist, our view is monist in that both mind and body ensue from particular processes and configurations in fundamental space-time geometry. Oxford philosopher Galen Strawson (unpublished) describes such a view as 'realistic monism'

How did we arrive at this conclusion? Roger Penrose's non-compatibility was the clue, a thread with which to unravel larger mysteries, such as the nature of experience and free will. Penrose followed this thread, seeking a non-computable physical process that could occur in the brain. He nominated as a candidate a particular type of self-organizing collapse of the quantum wave function ('objective reduction' OR[4-7]).

In the Penrose view, quantum superpositions (e.g. two alternate states or locations of an object existing simultaneously) are actually slight separations ('bubbles' in the underlying make-up of reality (space-time geometry). If isolated and thus able to persist, the separations become unstable and eventually reach a threshold collapse (this instability in space-time separation is the link to quantum gravity). At the instant of collapse each space-time bubble reduces to a definite, unseparated state as an OR event occurs. In each event the choice of sta is selected, non-computably, to reflect some influence that is neither random nor completely deterministic, but c to hidden propensities embedded in fundamental space-time. A series of such events may be seen as a pattern o bubbles and ripples at the smallest scale in the make-up of reality.

How does this relate to the problem of experience? Panpsychist and panexperiential philosophers have been claiming for thousands of years that varieties of proto-conscious experience ('qualia' are intrinsic properties of reality. Qualia might be particular patterns in fundamental space-time geometry, for example, patterns encoded Planck-scale spin networks. OR events might occur in and of an experiential medium.

Regarding free will, the problem is that our actions seem neither totally deterministic nor random (probabilistic) The only other apparent choice is Penrose's non-computability. In the Orch OR model, microtubule quantum superpositions compute and evolve linearly (analogous to a quantum computer) during pre-conscious processin but are influenced at the instant of OR collapse by hidden (Platonic) non-computable logic inherent in space-tim geometry. The precise outcome–our free will actions–results from effects of the hidden logic on the quantum system poised at the edge of objective reduction.

As an illustration consider a sailboard in which a sailor sets the sail in a certain way; the direction in which the board sails is determined by the action of the wind on the sail (Fig. 1). Imagine that the sailor is a non-consciou

robot/zombie run by a quantum computer that is trained and programmed to sail. Setting and adjusting of the sa sensing the wind and so forth are algorithmic and deterministic, and analogous to the pre-conscious, quantum computing phase of Orch OR. Direction and intensity of the wind (seemingly capricious or unpredictable) are analogous to hidden, non-local variables (e.g. 'Platonic'quantum-mathematical logic inherent in space-time geometry). The choice, or outcome (the direction the board sails, the point on shore where it lands) depends on deterministic sail settings acted on repeatedly by the apparently unpredictable wind. In a similar way, our 'free will'actions could be the net results of deterministic processes acted on by hidden quantum logic at each Orch C event.

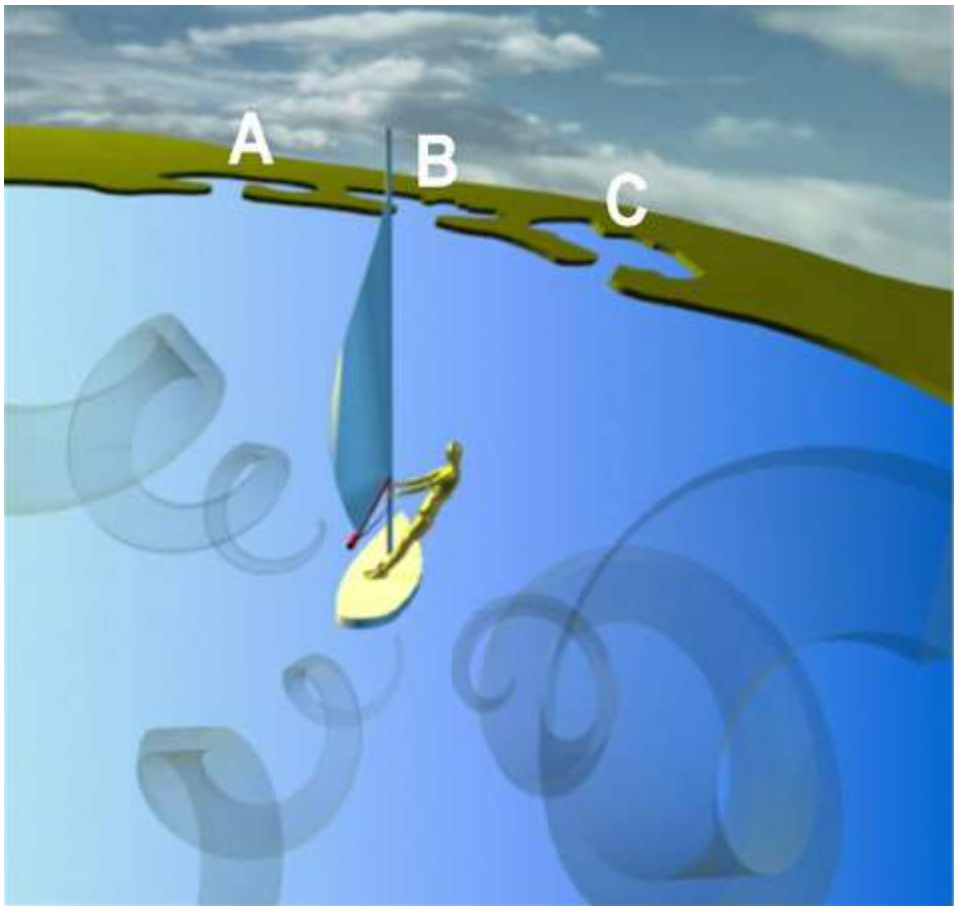

Figure 1. Free will may be seen as the result of deterministic processes (behavior of trained robot windsurfer) a on repeatedly by non-computable influences, here represented as a seemingly capricious wind.

To this point Spier and Thomas raise an appropriate objection: 'because there must, by definition, be only one Platonic logic, then all agents performing under it must hold the same opinion' There are several reasons for thi not to be the case. First, Planck-scale geometry might be evolving and thus variable over time[8]. Secondly, its influence would still be subject to the vagaries of individual personalities and potentially overriding determinist processes. Consequently, reactions might vary, owing to genetic influences, learned behavior and intensity of deterministic drives. Thirdly, effects of the proposed hidden propensities (on protein conformation) are subtle,

requiring a properly susceptible 'frame of mind'

I completely agree with Spier and Thomas that protein molecules 'act as the basic computational elements in liv cells' but simply add that proteins could act as quantum-computational elements. Proteins are dynamic–their sta delicate balance between various countervailing forces. However, strong forces cancel out so that dynamic prot conformation and function are determined by weak quantum-level dipole interactions called London (van der Waals) forces[9]. These forces occur largely in intra-protein hydrophobic pockets; water-excluding regions at whi anesthetic molecules act[10]. If proteins are quantum switches, then an interactive protein lattice (e.g. microtubule could constitute a quantum computer.

Spier and Thomas also argue that 'microtubules are too unstable to account for consciousness' While this is true non-neuronal dividing cells, whose microtubules radiate from centrioles and are (as Spier and Thomas describe) dynamically unstable, neurons in the brain don't divide, their microtubules do not radiate from centrioles, and th do not manifest dynamic instability[11]. Brain microtubules are quite stable and interlinked in complex cytoskelet networks.

Is the single-celled animal paramecium conscious? Spier and Thomas assume that the answer from the Orch OR position is yes, because we describe intelligent paramecium behavior mediated by microtubules. In fact, we do n claim that a paramecium is conscious. Based on an upper limit of hundreds of milliseconds of sustained quantum coherence, the Orch OR model predicts a lower limit for consciousness at the level of about 300 neurons (e.g. small worms and urchins). A single-celled paramecium, while clever, seems unlikely to sustain sufficient quant coherence to reach threshold for OR reduction (up to one minute would be required), and is thus unlikely to atta conscious experience[11]. One may question even rudimentary consciousness in small worms ("What is it like to b worm?") but, unlike any other theory, Orch OR is at least able to make such a prediction.

What about even more primitive cells? Spier and Thomas observe that bacteria, which also seem computational driven, lack microtubules, questioning their necessity for intracellular information processing. However, such bacteria do have 'protein-based circuits' and recent evidence shows the structure of such proteins to be strikingly similar to the microtubule protein, tubulin'SUP>13. Bacteria thus have primitive forms of microtubules. Althou Orch OR predicts emergence of consciousness at roughly 300 neurons, pre-conscious protein-based quantum computation might be an essential feature of all living protoplasm.

Finally, Spier and Thomas ask what it is in Orch OR that 'does the work of the mind' As in other theories, the answer is mainly glucose and oxygen. We discard nothing from conventional theories except the assumption tha consciousness emerges completely from membrane-level computational complexity. If one views a particular conscious experience or volitional choice as correlating with a particular neural network setting into specific attractor dynamics, one may continue to do so, and simply add underlying regulation by quantum computation i microtubules. Orch OR is a 'funds-mental'extension of neural-level theories of consciousness.

References

1. Hameroff, S. (1998) Funds-mental geometry: The Penrose-Hameroff Orch OR model of consciousness, Geometry and the Foundations of Science: Contributions from Oxford Conference Honoring Roger Penrose (Huggett, S. et al., eds) pp 103-127, Oxford University Press.
2. Hameroff, S. (1998) 'More neural than thou' reply to Churchland's Brainshy, in Toward a Science of Consciousness II: The Second Tucson Discussions and Debates (Hameroff, S., Kaszniak, A., and Scott,

eds), pp. 197-213, MIT Press.
3. Hameroff, S. (1998) Did consciousness cause the Cambrian evolutionary explosion?, in Toward a Scien of Consciousness II: The Second Tucson Discussions and Debates (Hameroff, S., Kaszniak, A. and Scot A., eds), pp. 421-437, MIT press.
4. Penrose, R. (1989) The Emperor's New Mind, Oxford University Press.
5. Penrose, R. (1994) Shadows of the Mind, Oxford University Press.
6. Penrose, R. (1996) On gravity's role in quantum state reduction Gen. Relativity Gravitation 29, 551-600
7. Penrose, R. (1997) On understanding understanding Int Stud. Philos. Sci. 11, 7-20. Smolin, L. (1997) Li of the Cosmos, Oxford University Press. Voet, D. and Voet, J.G. (1995) Biochemistry (2nd edn), Wiley Franks, N.P. and Lieb, W.R. (1982) Molecular mechanisms of general anesthesia. Nature 316, 349-351. Penrose, R. and Hameroff, S.R. (1995) What gaps? Reply to Grush and Churchland J. Conscious. Stud. 99-112. Hameroff, S.R. and Penrose, R. (1996) Conscious events as orchestrated space-time selections. Conscious. Stud. 3, 36-53.
8. Lowe, J. and Amos, L.A. (1998) Crystal structure of the bacterial cell-division protein FtsZ. Nature 391 203-206.

Reply to Hameroff from Emmet Spier and Adrian Thomas

We thank Stuart Hameroff for detailing the intellectual history of the Orch OR model. However, re-rooting its foundations in a form of monism begs more questions than it answers. Even ignoring the problematic conclusio that ' then 'all forms of matter (rocks, air, grass, protozoa, etc.) have minds, we still find it hard to understand ho his mind-component of matter influences the substance component of matter. Either its influence is local–in wh case 'learned behaviour'and 'intensity of deterministic drives'would be neural juggernauts affected by rain drops global, in which case our previous, single Platonic-mind criticism still stands.

Neither are we convinced that the mechanistic side of the brain is a computer, even less a quantum computer. A computer manipulates memory elements. Merely identifying the memory elements (qubits) as 'microtubule quantum superpositions' completely ignores the fact that machinery must exist for these bits to be fetched, processed and re-stored. This is the job of the central processing unit (CPU) and is wholly ignored in models of brain as a computer. We also note that there is nothing particularly useful about being a quantum computer, unl factorizing large numbers'(their one established novel capacity) is a fundamental process of mind.

Hameroff agrees with Bray's conclusions that protein molecules are the basic operational elements in living cel but goes on to suggest that quantum mechanics might play some role too. Bray[2] has clearly demonstrated that chemistry is sufficient to explain the complex behaviour of single-celled organisms such as *E. coli*. Furthermore mutations that alter the chemistry of *E. coli's* proteins change its behaviour in ways that are predictable from th effect of the mutations on protein chemistry. Conventional chemical mechanisms can explain how single cells process such information 'invoking the Orch OR model is neither necessary nor justified.

We also note that Hameroff has suggested that the emergence of consciousness in primitive metazoans (worms) might explain the Cambrian explosion. However, recent fossil finds have reduced the Cambrian anomaly. Paper describe 570 million-year old (Precambrian) fossils of metazoan embryos; such fossil finds establish that metazoans originated long before the Cambrian explosions, but these Precambrian animals did not fossilize.

Crucial to the Orch OR model is the idea that 'brain microtubules are quite stable'and that they can maintain son thermal isolation from the rest of the body. This is a misconception. Microtubules in the axons of neurons (but n

in the dendrites or cell body) are indeed held in a stable configuration; but even in axons the microtubules turn over at high rates and being shorter than the axon their open ends are exposed to the cell contents[6,7]. The fluid-filled lumen of microtubules can act as a transport route-for example, during bacterial sex, DNA is transported down the center of a microtubule-like protein1. It seems unlikely, when molecules as large as DNA, let alone water, can flow through microtubules, that they could possibly maintain the thermal isolation that is an essentia requirement for the Orch OR model.

Indeed, 'how many roads must a man walk down before you call him a man?" possibly, according to the Orch O model, 'the answer [might well be] blowing in the wind[8].

References:

1. Shor, P. (1997) Polynomial-time algorithms for prime factorization and discrete logarithms on a quantu computer. SIAM J. Comp. 26, 1484-1509. Bray, D. (1995) Protein molecules as computational element living cells. Nature 376, 307-312. Xiao, S., Zhang, Y. and Knoll, A.H. (1998) Three-dimensional preservation of algae and animal embryos in a neoproterozoic phosphorite. Nature 391, 553-558. Li, C-W Chen, J-Y. and Hua, T-E. (1998) Precambrian sponges with cellular structures. Science 279, 879-882. Thomas, A.L.R. (1997) The breath of life: Did increased oxygen levels trigger the Cambrian Explosion? Trends Ecol. Evol. 128, 44-45. Bray, D. (1992) Cell Movements, Garland Publishing. Slaughter, T., Wa J. and Black, M.M. (1997) Microtubule transport from the cell body into the axons of growing neurons. Neuroscience 17, 5807-5819.
2. Dylan, B. (1962) Blowing in the Wind, Warner Bros.