What 'Gaps'?
Reply to Grush and Churchland

Journal of Consciousness Studies, 2, No. 2, 1995, pp. 99-112

Roger Penrose, Oxford University, Mathematical Institute
24 29 St. Giles, Oxford OXl 3LB, UK

Stuart Hameroff, Departments of Anesthesiology and Psychology
Associate Director, Center for Consciousness Studies
The University of Arizona, Tucson, AZ 85724, USA.
email: hameroff@u.arizona.edu

Abstract

Grush and Churchland (1995) attempt to address aspects of the proposal that we have been making concerning a possible physical mechanism underlying the phenomenon of consciousness. Unfortunately, they employ arguments that are highly misleading and, in some important respects, factually incorrect. Their article ‘Gaps in Penrose’s Toilings’ is addressed specifically at the writings of one of us (Penrose), but since the particular model they attack is one put forward by both of us (Hameroff and Penrose, 1995; 1996), it is appropriate that we both reply; but since our individual remarks refer to different aspects of their criticism we are commenting on their article separately. The logical arguments discussed by Grush and Churchland, and the related physics are answered in Part l by Penrose, largely by pointing out precisely where these arguments have already been treated in detail in Shadows of the Mind (Penrose, 1994). In Part 2, Hameroff replies to various points on the biological side, showing for example how they have seriously misunderstood what they refer to as ‘physiological evidence’ regarding to effects of the drug colchicine. The reply serves also to discuss aspects of our model ‘orchestrated objective reduction in brain microtubules – Orch OR’ which attempts to deal with the serious problems of consciousness more directly and completely than any previous theory.

Part 1: The Relevance of Logic and Physics

Logical arguments

It has been argued in the books by one of us, The Emperor’s New Mind (Penrose, 1989 – henceforth Emperor) and Shadows of the Mind (Penrose, 1994 – henceforth Shadows) that Gödel’s theorem shows that there must be something non–computational involved in mathematical thinking. The Grush and Churchland (1995 – henceforth G&C) discussion attempts to dismiss this argument from Gödel’s theorem on certain grounds. However, the main points that they put forward are ones which have been amply addressed in Shadows. It is very hard to understand how G&C can make the claims that they do without giving any indication that virtually all their points are explicitly taken into account in Shadows. It might be the case that the arguments given in Shadows are in some respects inadequate, and it would have been interesting if G&C had provided a detailed commentary on these particular arguments, pointing out possible shortcomings where they occur. But it would seem from what G&C say that they have not even read, and certainly not understood, these arguments. A natural reaction to their commentary would be simply to say "go and read the book and come back when you have understood its arguments."[1] However, it will be helpful to pinpoint the specific issues that they raise here, and to point out the places in Shadows where these issues are addressed.

The main argument that they appear to be raising against Penrose's (1989; 1994) use of Gödel's theorem (to demonstrate non–computability in mathematical thinking) is that mathematical thinking contains errors. They give the impression that the possibility of errors by mathematicians is not even considered by Penrose. However, in §§3.2, 3.4, 3.17, 3.19, 3.20 and 3.21 of Shadows the question of possible errors in human or robot mathematical reasoning is explicitly addressed at length. (The words 'errors' and 'erroneous' even appear explicitly in the headings of two of those sections and it is hard to see why G&C make no reference to these parts of the book.) In addition, on page 16 of their commentary, G&C claim that 'most of the technical machinery' involved in Penrose's arguments refer to what they call 'Ala' and 'Alc', on their page 15, which they choose not to dispute; whereas in fact by far the most difficult technical arguments given in Shadows are those which specifically address the possibility of errors in human or robot mathematical reasoning (these are given in §§ 3.19 and 3.20 of Shadows). It is difficult to understand why G&C fail to refer to this discussion, seeming to suggest (quite incorrectly) that Penrose has an in-built faith in the complete accuracy in the reasoning of mathematicians! G&C have a curious way of formalizing what they believe to be the ingredients of Penrose's arguments. In particular, on page 16 they refer to 'Penrose's Premise A 1: Human thought, at least in some instances, perhaps in all, is sound, yet non-algorithmic' (which they break down into A I a, . . . , A 1 e). Their 'Premise A 1' is nowhere to be found in Penrose's writings. It is fully admitted by Penrose that actual human thinking can be unsound even when seeming to be carried out in the most rigorous fashion by mathematicians.

It may well be that there is a genuine and deep misunderstanding implicit in what G&C are attempting to say, and it may be helpful to try to clarify the issue here. For the purposes of our present discussion (and for the essential discussion given in Shadows) it will be sufficient to restrict attention to a very specific class of mathematical statements, namely those referred to as "pi 1"–sentences. Such sentences are assertions that particular (Turing–machine) computations do not halt.

There are some very famous examples of mathematical assertions which take the form of "pi 1"–sentences, the best known being the so–called 'Fermat's Last Theorem'. Other examples are 'Goldbach's conjecture' (still unproved) that every even number greater than 2 is the sum of two primes, Lagrange's Theorem that every natural number is the sum of four squares, and the famous 4–colour theorem. It is useful to concentrate one's attention on "pi 1"–sentences because this is all one needs for application of the Gödel argument to the issue of computability in human mathematical thinking. There is no relevant issue of dispute between mathematicians as to the meaningfulness and objectivity of the truth of such sentences. (One might, however, worry about the 'intuitionists' or other constructivists in this context –and some reference to such viewpoints is given on p. 18 and footnote 30 on p. 20 of the G&C article. However, such constructivist viewpoints do not evade the Gödel argument and the use made of it in Shadows as is explicitly addressed in the discussion of Q9 on page 87 of Shadows, a discussion not even referred to by G&C.)

As far as we can make out, G&C are not disputing the absolute (‘Platonic’) nature of the truth or falsity of explicit "pi 1"–sentences. The issue is the accessibility of the truth of "pi 1"–sentences by human reasoning and insight.

We should make clear what is meant by a word such as ‘accessibility’ in this context, since there seem to be a great many misconceptions by philosophers and others as to how mathematical understanding actually operates. It is not a question of some kind of ‘mystical intuition’ that (some) mathematicians might have, and which is unavailable to ordinary mortals. What is being referred to by ‘access’ is simply the normal procedure of mathematical proof. It is not even a question of how some mathematician might have the inspiration to arrive at a proof. It is merely the question of the understanding which is involved in the ability to follow a proof in principle. (See, in particular, in the response to Q12 pp. 101 3 of Shadows.) However, it should be made clear in this context that the word ‘proof’ does not refer necessarily to a formalized argument within some pre-assigned logical scheme. For example, the arguments given by Andrew Wiles (as completed by Taylor and Wiles) to demonstrate the validity of Fermat’s last theorem were certainly not presented as formal arguments, within, say, the Zermelo–Fraenkel axiom system. The essential point about such arguments is that they have to be correct as mathematical reasoning. It is a secondary matter to try to find out within which formal mathematical systems such arguments can be formulated. Indeed, what the Gödel argument shows (and this is not in dispute) is that if the rules of some formal system, F, can be trusted as providing correct demonstrations of mathematical statements —and here we need restrict attention only to "pi 1"–sentences—then the particular "pi 1"–sentence G(F) must also be accepted as true even though it is not a consequence of the very rules provided by F. (Here the sentence G(F) is the Gödel proposition which asserts the consistency of the formal system F—assuming that F is sufficiently extensive. It can also be taken as the explicit statement $C_k(k)$ exhibited on p. 75 of Shadows.[2] What this shows is that mathematical understanding (i.e. mathematical proof-in the sense above) cannot be encapsulated in any humanly acceptable formal system. Here ‘acceptable’ means acceptable to mathematicians as a reliable means of obtaining mathematical truths, where attention may be restricted to the truth of "pi 1"–sentences.

The notion of ‘proof’ that is being referred to above certainly raises profound issues. However, it would be unreasonable to dismiss it as something which is too ill–defined for scientific consideration or perhaps ‘mystical’. There is indeed something mysterious about the very nature of ‘understanding’ and this is what is involved here. But the notion of proof that is involved in mathematical understanding is extraordinarily precise and accurate. There is no other form of argument within science or philosophy which really bears comparison with it. Moreover, this notion transcends any individual mathematician. But it is what mathematicians individually strive for. If one mathematician claims to have an argument for demonstrating the validity of some assertion—say a "pi 1"–sentence—then it should in principle be possible to convince another mathematician that the argument, and hence the conclusion, is correct - unless there is an error, in which case it is up to the mathematicians to locate this error. There is no question but that mathematicians do, not infrequently, make errors. This is not the point. The point is that it is possible for there actually to be an argument of some nature, to be found, which demonstrates the truth of the "pi 1"–sentence in question, and it is the mathematicians' business to try to find such an argument, whether or not they make mistakes in the process of attempting to achieve this. If there is an argument accessible to human understanding, then 'access' to this particular "pi 1"–sentence is possible. The point about the Gödel argument in this context is that there is no way to encapsulate the means whereby this access is achieved, in terms of mathematically acceptable computational rules.

It is no part of Penrose's argument against computationalism that al! "pi 1"–sentences should be humanly accessible (although there is some discussion of this possibility in chapter 8 of Shadows). What is argued for in Shadows is the assertion that the class of "pi 1"–sentences which are in principle humanly accessible is not a class which is computationally accessible (technically; not a recursive set—it is certainly not a knowably recursive set).

The issue of practical accessibility (as opposed to in principle accessibility) of "pi 1"–sentences by individual mathematicians, or perhaps by the mathematical community as a whole, is of course a somewhat separate matter, but these issues are discussed in considerable detail in Chapter 3 of Shadows (and also in the discussion of Q8 on pp. 83J9). No mention of this is made by G&C. Nor do they refer to the fact that in Shadows the Gödel argument is applied to computability both in the 'in principle' and the 'in practice' sense, according to context.

A detailed commentary on all of the remaining misunderstandings of Penrose's arguments displayed by G&C would take more space than is available to us here. However, it will be helpful to list some of the more important of these. At the end of the paragraph which finishes at the top of page 17, in a rather confused sentence, they seem to be saying that there are no sound procedures of mathematical deliberation, beyond, say, Zermelo–Fraenkel set theory (ZF). This shows that they do not understand Gödel. He demonstrated that if ZF is sound, then G(ZF) does indeed enable one to go beyond ZF (etc.). On page 18 they refer to the changing perceptions of mathematical rigour over the centuries, seeming to suggest that Cauchy and Cantor might have had different conceptions of what is unassailably true in mathematics. (Here, 'unassailable' refers merely to 'correctly proved' in the sense given above.) There is nothing wrong with the continual broadening of the kind of reasoning which can be used to obtain mathematical results, say, "pi 1"–sentences, and what Cantor did was to increase this breadth enormously over what had been achieved before. This led to delicate, and sometimes disputed issues, such as the Axiom of Choice. The possibility of differing viewpoints with regard to the Axiom of Choice is explicitly addressed in the discussion of Q11 on pp. 97-101 of Shadows (although G&C seem oblivious of this fact). The conclusion is that this does not significantly affect the non–computability argument.

G&C make a number of points which seem to suggest that they think that Penrose is unaware of certain elementary logical points. This is particularly irritating. For example at the bottom of p.19 they point out that: 'M does not believe that A is unassailably true' does not entail that 'M disbelieves A'. Of course it does not. This kind of distinction is essential to the arguments of Chapter 3 of Shadows.

Physical arguments

There is a basic confusion in G&C p. 15, concerning the algorithmic or non–algorithmic nature of physical laws. The issue of approximations is brought out in Shadows only as a stop-gap measure to handle the continuous parameters in terms of which modem physical theories are invariably described. Artificial neural networks (ANNs) are simply computational devices, being digitally, computationally controlled, and there is no issue of approximation involved. Moreover, there is surely no serious suggestion that the present–day ANNs are actually conscious, or possess genuine understanding, despite all their 'successes'. The Penrose argument implies that we must go beyond such computational action if we are to find a physical basis for consciousness. ANNs are extensively considered in Shadows, and constitute an important part of the arguments of Chapter 3.

G&C spend some time in their article attacking the suggestion that the growth of quasicrystals might be non–computational. This is a red herring. It should be made clear that such a suggestion is nowhere made in Penrose's writings. In Emperor (but not in Shadows) quasicrystals are mentioned, and the suggestion is made that there might be something essentially non–local (not non–computational) in quasicrystal assembly which could perhaps involve the unknown physics underlying quantum state reduction in an essential way.

G&C claim that there is no physical indication that the unknown theory of 'quantum gravity' (which correctly combines gravitational theory with quantum mechanics) should be non–computational. Penrose does not make any strong claim for such noncomputability from purely physical theory. However, despite what G&C say, such evidence does exist and is discussed in §§7.8 and 7.10 in Shadows.

Finally, we come to what G&C call the 'convenient myth' of Platonism (p.21), where they refer to Fig 8.1 on p. 414 of Shadows. The main worry that they (and others) seem to have about Platonism relates to 'mystery 3' of that diagram, namely the relationship between human thought and absolute mathematical truth.

However, the real mystery for mathematical physicists is 'mystery 1': how is it that the physical world indeed accords—and has accorded since the beginning of time—with such extraordinary precision with subtle and beautiful mathematical laws. If these mathematical laws are merely the product of our recent mental activity, then we are presented with a profound paradox.

Part 2: The Biological Side

Remarks by G&C pertaining to microtubules (MTs), quantum theory and consciousness fall into the following areas:

1. How could quantum gravity, MTs and consciousness 'conceivably' be inter–related?
2. How could consciousness depend on MTs when 'physiological evidence demonstrates' that consciousness can occur without them?
3. How do MTs communicate with neural membrane and synaptic functions? How can MTs encode and process information?
4. Given the apparently noisy, thermal environment within neurons and the brain, how could quantum coherent phenomena occur a) within neurons, and b) throughout macroscopic brain regions?

5. Why don’t ions such as sodium and calcium prevent quantum phenomena in cytoplasm? Within hollow MT cores?
6. Miscellaneous

1. How could quantum gravity, MTs and consciousness ‘conceivably’ be inter–related?

Since papers describing our model of ‘orchestrated objective reduction in brain microtubules—Orch OR’ (Hameroff and Penrose, 1995; 1996; forthcoming) had not been published, nor reviewed by G&C at the time their critique was written, we summarize key points of the model here:

(1) Aspects of quantum theory (e.g. quantum coherence) and of the suggested physical phenomenon of quantum wave function ‘self–collapse’ (objective reduction: OR – Penrose, 1994) are essential for consciousness, and occur in cytoskeletal microtubules (MTs) and other structures within each of the brain’s neurons.

(2) Conformational states of MT subunits (tubulins) are coupled to internal quantum events, and cooperatively interact with other tubulins in both classical and quantum computation.

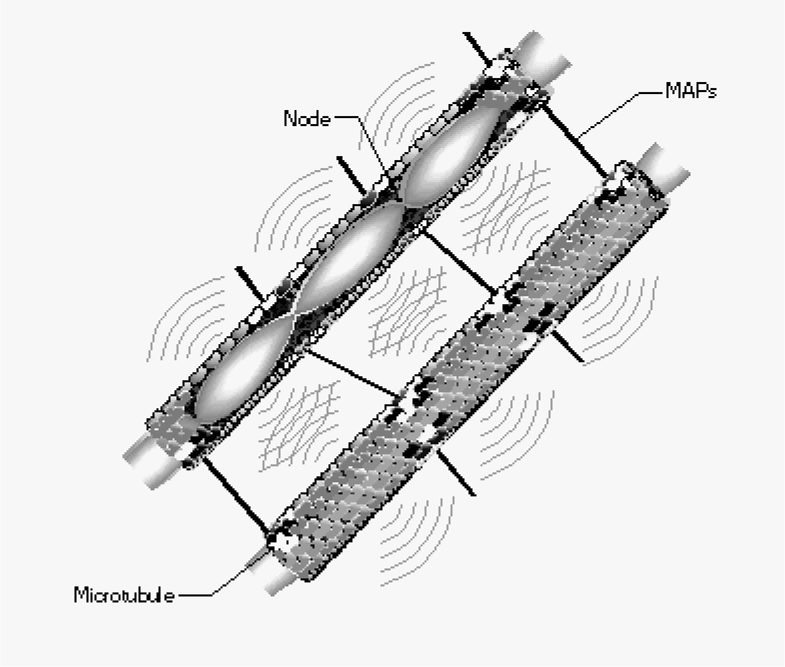

Figure 1. *Schematic representation of proposed quantum coherence in microtubules. Having emerged from resonance in classical automaton patterns, quantum coherence non-locally links superpositioned tubulins (grey) within and among microtubules. Upper microtubule: cutaway view shows coherent photons generated by quantum ordering of water on tubulin surfaces, propagating in microtubule waveguide. MAP (microtubule–associated–protein) attachments breach isolation and prevent quantum coherence; MAP attachment sites thus act as ‘nodes’ which tune and orchestrate quantum oscillations and set possibilities and probabilities for collapse outcomes (‘orchestrated objective reduction – Orch OR’).*

(3) Quantum coherence occurs among tubulins in MTs, pumped by thermal and biochemical energies (e.g. Fröhlich, 1968; 1970; 1975). Water at MT surfaces is 'ordered' – dynamically coupled to the protein surface. Water ordering within the hollow MT core (acting like a quantum waveguide) can result in quantum coherent photons ('super-radiance' and 'self–induced transparency' –Jibu et al., 1994; 1995).

(4) During pre–conscious processing, quantum coherent superposition/computation occurs in MT tubulins and continues until the mass-distribution difference among the separated states of tubulins reaches a threshold related to quantum gravity. Self-collapse (OR) then occurs.

(5) The OR self-collapse process results in classical 'outcome states' of MT tubulins which then implement neurophysiological functions. According to principles of OR (Penrose, 1994), the outcome states are 'non–computable'; that is they cannot be determined algorithmically from the tubulin states at the beginning of the quantum computation.

(6) Possibilities and probabilities for post–OR tubulin states are influenced by factors including initial tubulin states, and attachments of microtubule–associated proteins (MAPs) acting as 'nodes' which tune and 'orchestrate' the quantum oscillations (Figure 1). We thus term the self-tuning OR process in microtubules 'orchestrated objective reduction – Orch OR'.

(7) According to the arguments for OR put forth in Penrose (1994), superpositioned states each have their own spacetime geometries. When the degree of coherent mass energy difference leads to sufficient separation of spacetime geometry, the system must choose and decay (reduce, collapse) to a single universe state. Thus Orch OR involves self–selections in fundamental spacetime geometry (Hameroff and Penrose, forthcoming).

(8) To quantify the Orch OR process, we calculate from the indeterminacy principle E=h/T, in which E is the gravitational self-energy of the quantum superposition (a mass acting on its displaced self), h(actually "hbar") is Planck's constant over 2p , and T is the coherence time (how long the coherence is sustained). If we assume a coherence time T = 500 msec (shown by Libet et al., 1979, and others to be a relevant time for pre-conscious processing), we calculate E, and determine the number of MT tubulins whose coherent superposition for 500 msec will elicit Orch OR. This turns out to be about 10^9 tubulins.

(9) A typical brain neuron has roughly 10^7 tubulins (Yu and Baas, 1994). If, say, 10 percent of tubulins within each neuron are involved in the quantum coherent state, then roughly 10^3 (one thousand) neurons would be required to sustain coherence for 500 msec, at which time the quantum gravity threshold is reached and Orch OR then occurs.

(10) We consider each self–organized Orch OR as a single conscious event; a series of such events would constitute a 'stream' of consciousness. If we assume some form of. excitatory input (e.g. you are threatened, or enchanted) in which quantum coherence emerges faster, then, for example, 10^{10} coherent tubulins could Orch OR after 50 msec, or 10^{11} after 5 msec. Turning to see a Bengal tiger in your face might perhaps elicit 10^{12} in 0.5 msec, or more tubulins, faster.[3] A slow emergence of coherence (your forgotten phone bill) may require longer times.

(11) Each instantaneous Orch OR may then ‘bind’ various superpositions which may have evolved in separated spatial distributions and over different time scales, but whose net displacement self–energy reaches threshold at a particular moment. Information is bound into an instantaneous event (a ‘conscious now’). Cascades of Orch ORs define a forward flow, or stream of consciousness.

The Orch OR model thus accommodates some important features of consciousness:
1) control/regulation of neural action, 2) pre-conscious to conscious transition, 3) noncomputability, 4) causality, 5) binding of various (time scale and spatial) superpositions into instantaneous ‘now’, 6) a ‘flow’ of time, and 7) a connection to fundamental space-time geometry.

2. How could consciousness depend on MTs when physiological evidence demonstrates’ that consciousness can occur without them? The ‘physiological evidence’ provided by G&C turns out to be:

1. the drug colchicine causes depolymerization of MTs
2. patients taking colchicine pills for treatment of gout, and experimental animals receiving colchicine directly to the brain or cerebrospinal fluid don’t become unconscious. (G&C fail to mention that the animals do become ‘demented’.)

The drug colchicine (from the plant Colchicum automnale) has been known as a poison since antiquity, and used as medicine since the 18th century (Dustin, 1984). It is still used for the treatment of gout, first described by Hippocrates (fifth century BC) and caused by accumulation of urate crystals in joint spaces. Immune cells (lymphocytes and macrophages) migrate to, and engulf the ‘foreign’ crystals. The cells release inflammatory substances which cause the joint swelling and extreme pain characteristic of gout. Cell migration and other activities depend on dynamic rearrangements of their cytoskeletons, including, and depending primarily on, MTs. For example in locomotion, MTs inside the cells self–assemble in the proper orientation; other cell components then follow or are transported. Colchicine prevents the immune cell locomotion and other activities by binding to growing (‘beta plus’) ends of actively polymerizing MTs. This prevents the cycles of MT polymerization and depolymerizeration required during cell division (MT ‘mitotic spindles’), and locomotion and other immune cell activities responsible for symptoms of gout and other inflammatory processes. In patients taking colchicine pills for gout, drug access to the brain through the blood–brain barrier is about $l0^{-4}$ of blood levels (Bennett et al., 1981). So it’s not surprising in any case that these patients have no central nervous system symptoms.

G&C cite studies (e.g. Bensimon and Chemat, 1991; Kolasa et al., 1992; Emerich and Walsh, 1991) in which colchicine delivered directly to the brain or cerebrospinal fluid of experimental animals failed to cause loss of consciousness. G&C neglect to mention that the colchicine does cause significant cognitive impairments of learning and memory. Bensimon and Chernat (1991), for example, characterized a ‘dementia’ in their colchicine–treated animals which they likened to Alzheimer’s disease (often linked to MT disruption – e.g. Matsuyama and Jarvik, 1992).

G&C are thus asking: if colchicine depolymerizes MTs, why doesn’t it cause unconsciousness (rather than just dementia)?Answer: even when given in huge doses directly to the brain, colchicine has minimal depolymerizing effects on brain MTs.Unlike the labile MTs in cell division and locomotion, most MTs in brain neurons are stable, ‘hardened’ by biochemical changes, do not engage in cycles of polymerization and depolymerization, and are resistant to colchicine. They have no exposed ‘beta plus ends’!Dynamic neuronal MTs which do undergo repeating cycles of polymerization and depolymerization are those involved in restructuring synaptic connections, perhaps accounting for colchicine’s impairments of learning and memory.

3. How do MTs communicate with neural membrane and synaptic functions? How can MTs encode and process information?

This topic is discussed in great detail in Hameroff (1987), Rasmussen et al. (1990), Dayhoff et al. (1994) and Hameroff and Penrose (1996).

As a general summary, MTs are linked to membrane receptors and ion channels both structurally, via smaller cytoskeletal proteins like fodrin, actin, synapsin and others, and biochemically, for example as part of second messenger cascades. MT actions define and modify neural architecture and synaptic connections and strengths. The MT cytoskeleton is each neuron’s nervous system.

As one example, consider ‘MAP–2’, a dendrite–specific, MT-crosslinking ‘microtubule–associated–protein’ (‘MAP’), necessary for learning and memory. As a result of synaptic membrane receptor activation (e.g. Halpain and Greengard, 1990), MAP–2 is ‘dephosphorylated’ (imparting energy and information to the cytoskeleton). This process is essential to strengthening synaptic pathways, for example in cat visual cortex with visual stimulation (Aoki and Siekevitz, 1985) and rat temporal cortex in auditory Pavlovian conditioning (Woolf et al., 1994). MAP–2 dephosphorylation, which consumes a large proportion of brain biochemical energy (e.g. Theurkauf and Vallee, 1983) acts to reconfigure the sub–synaptic cytoskeleton (Bigot and Hunt, 1990; Friedrich, 1990). Regarding neurotransmitter release, it is true as G&C say that MTs don’t extend into the terminal axon region (nor into dendritic spines – MTs are too large). MTs do, however, connect to smaller cytoskeletal proteins like synapsin which, influenced by calcium ion fluxes, are directly involved in neurotransmitter release (Hirokawa, 1991). (In dendrites, MTs connect to smaller actin filaments which extend into dendritic spines and interact with receptors.) As Beck and Eccles (1992) point out, the process of neurotransmitter release has a seemingly random, probabilistic component (only about one sixth of axonal depolarizations result in neurotransmitter vesicle release). Beck and Eccles suggest this may reflect some unrecognized quantum influence, although we don’t share their view that it is a purely ‘dualist’ influence.

How can MTs encode, and process information? Vassilev et al. (1985) demonstrated signal transmission along tubulin chains. As cylindrical lattices of tubulin dipoles, MTs are well suited to process information, and a number of models of MT signalling and information processing have been suggested. These include propagating tubulin conformational changes (Atema, 1973; Roth and Pihlaja, 1977; Hameroff and Watt, 1982), sequential phosphorylation/dephosphorylation along MT tubulins (Puck and Krystosek, 1992), tensegrity (Wang and lngber, 1994), non–linear soliton waves (Chou et al., 1994; Sataric et al., 1992), spinglass and ferroelectric effects (Tuszynski et al., 1995), ‘cellular automaton’ behavior (e.g. Rasmussen et al., 1990) and quantum coherent photons (Jibu et al., 1994; 1995).

4. Given the apparently noisy, thermal environment within neurons and the brain, how could quantum coherent phenomena occur: a) within neurons, and b) throughout macroscopic brain regions?

This is the crux of the matter. How can quantum effects emerge and be sustained in or around neurons for physiological time durations? The cellular milieu is considered noisy, thermal and chaotic. Ionic fluxes such as axonal depolarization would seem to disrupt quantum coherent effects, if they did occur. We consider 3 types of quantum phenomena: **Quantum coherent superposition (‘Fröhlich mechanism’)** leading to self–collapse (Orch OR) among hydrophobic pockets in MT tubulins (Hameroff and Penrose, 1995;1996). **Dynamic ordering of water** on MT outer surfaces, in cytoplasm and extracellular spaces (Ricciardi and Umezawa, 1967; Del Giudice et al., 1983). **Coherent photons** (‘super–radiance and self–induced transparency’) in hollow MT ‘waveguides’ (Jibu et al, 1994; 1995).The internal cell environment in which MTs (and life) exist is the cytoplasm. The cytoplasm within neurons and the surrounding extracellular space differs markedly from non-living aqueous media, considered as water molecules in continuous thermal agitation. For example, proteins dissolved in the cytoplasm may be triggered by ionic fluxes to assemble rapidly into ‘gelatinous’ layers or structures. Cell cytoplasm may thus fluctuate between liquid solution (‘sol’), and more solid-state gelatin (‘gel’).Extensive charged surfaces on the cytoskeleton, membranes, organelles and extra-cellular matrices bind and order cytoplasmic water, which engage in cooperative dynamics. Several layers of ordered water on each of these many surfaces are predicted (e.g. Clegg, 1983), so that large proportions of cell interiors may be either dynamically ordered, gelatinous or solid.

According to Fröhlich(1968; 1970; 1975), dipole biomolecules structurally confined in membranes, MTs (and ordered water on their surfaces) become excited coherently by biochemical and thermal energy. The excitations reduce to a common frequency mode (10^9 to 10^{11} Hz), somewhat like the quantum phenomenon of a Bose–Einstein condensate (Anderson et al., 1995). In Bose–Einstein condensates like superconductors, coherence is attained by extreme cooling to remove thermal vibrations; in lasers, and in the Fröhlich model, the coherence derives from energy pumping.

Fröhlich coherence among (hydrophobic pockets within) MT subunits has been proposed as a basis for information processing via neighbor tubulin dipole interactions (e.g. as in a ‘cellular automaton’: Rasmussen et al, 1990). The coherent dynamics are viewed also to order water at MT outer and inner surfaces; the cytoplasm can transiently assume a quantum coherent state.Jibu et al. (1995) have recently examined Albrecht–Buehler’s (1992) experimental results in which single cells detect, orient and move toward weak red/infra–red light signals using their cytoskeleton. They explain the surpisingly efficient photon propagation through cytoplasm and extra–cellular fluid by dynamical water ordering within, and outside the cell. According to Jibu et al., the ordering behaves as a nonlinear coherent optical device – a ‘water laser’ – which enables lossless propagation of quantum coherent photons for a distance of 50 microns. Larger macroscopic quantum states may occur by coalescence of these regions, or by means such as quantum electromagnetic effects (Jibu et al., 1995).The Jibu et al. calculation of 50 microns implies quantum coherent regions with volumes of roughly 10^{-6} ml, or 10^{-9} litre. If one estimates brain volume of one litre, this fraction (10^{-9}) of total brain is about the same as our calculated 10^9 (out of $10^{18}$to 10^{19} total) brain tubulins, and could represent a fundamental unit of consciousness: the gravitational self–energy E, for a given coherence time (in this case, T= 500 msec). E, of course, will vary inversely with T, akin to the relationship between wavelength and frequency in the electromagnetic spectrum.In summary, MTs and associated water can shield or isolate quantum coherence/self collapse (Orch OR) from thermal disruption by several possible means:

1. Water in cells, or at least several water layers at MT and other surfaces, becomes dynamically ordered, and participates in quantum states. Hydrophobic pockets within each tubulin are shielded from water, and thermal energy condenses to a coherent state akin to a Bose–Einstein condensate, as Fröhlich(1968; 1970; 1975) has suggested. Quantum events within the hydrophobic pockets can influence protein conformation, and MT function. Calcium coupled sol–gel transformations form ‘gelatinous’ layers which transiently isolate MTs (Hameroff and Penrose, 1996).
2. The hollow MT inner core offers several theoretical advantages: it is sheltered, the particular quantum phenomena predicted to occur there (super–radiance and self–induced transparency) are quicker than thermal effects, and the cylinder can act as a quantum optical waveguide.

5. Why don’t ions such as sodium acrd calcium prevent quantum phenomena in cytoplasm? Within hollow MT cores?

Effects of ions on the dynamically ordered structure of water depend on the relative size of each ion compared to the water molecule H_2O (Ergin, 1983; Uedaira and Osaka, 1989; Jibu et al., 1995;). Ions whose radii are smaller than the H_2O radius (1.38 angstroms, or A) do not disturb the ordering. Sodium ion (radius 0.98A), calcium (1.OOÅ) and magnesium (0.72Å) can all embed between ordered H_2O molecules without disturbing them. Ions whose radii are close to that of water can take part in the dynamical geometry by replacing a water molecule. Potassium (1.33Å) is in this category. Ions larger than H_2O will disturb ordering. Chloride (1.81Å) should therefore disrupt ‘Type 2’ phenomena. (Infra-neuronal chloride concentration is extremely low, however, except during the terminal phase of an action potential.)

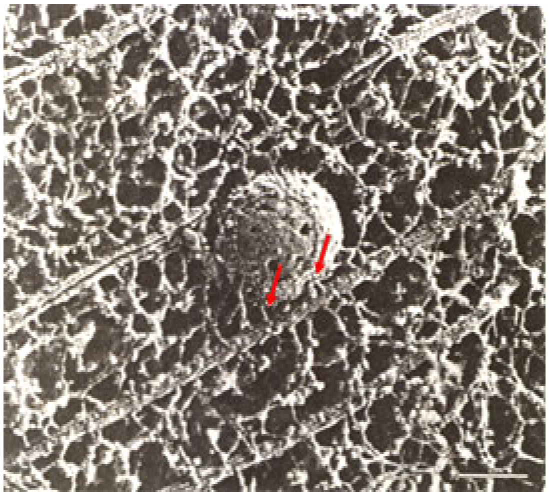

Figure 2. *Electron micrograph of axoplasmic transport of spherical vesicle by MTs. Approximately 4 MTs are visible, interconnected by MT–associated proteins (‘MAPS’). The MTs are arrayed in parallel from lower left to upper right. Arrows indicate short cross–bridges between vesicle and MT. Scale bar (lower right) is 100 nanometers. Reprinted with permission from Hirokawa, 1991.*

Within the hollow MT core, the ‘super–radiance and self-induced transparency’ predicted in Jibu et al. (1994) have characteristic times (to complete one cycle of photon emission and transfer) much shorter than the characteristic time of any thermally disordering phenomena.

In axons, chloride fluxes at the end of each action potential may ‘decohere’ Type 2 phenomena. In dendrites, graded potentials would seem to have less significant effects. We conclude that the quantum coherent superposition essential for consciousness occurs primarily in dendritic MTs. Recognition of dendrites as key regions related to consciousness is consistent with extensive work of Pribram (1991), Eccles (1992) and others.

6. Miscellaneous

G&C p. 23: ‘First we note that all cells have microtubules . . . ‘ [the implication presumably being ‘why isn’t your earlobe or some other bodily part conscious?’].

Penrose and Hameroff (P&H): Microtubules in neurons are quite distinct from those in other cells. 1) They are arrayed in parallel (rather than radially) because, unlike other cells, neurons lack centrioles. 2) They are quite stable. 3) They are far more abundant in neurons than other cells, form larger and complex networks, and have a greater genetic variability than other tissues (e.g. 17 isozymes of brain tubulin - Lee et al., 1986). 4) They perform specialized transport along axon and dendrite processes. (Figure 2) 5) They have neuron-specific MT– associated proteins (MAPs).

G&C p. 23: The ‘anaesthesia–microtubule connection.’

P&H: G&C imply that our view of the mechanism for general anaesthesia is depolymerization of MTs. This is incorrect. That idea emanated from Allison and Nunn (1968), in which the anaesthetic gas halothane depolymerized axopodia comprised of MTs in Actinosphaerium (but at 5 times the concentration required for clinical anaesthesia). Subsequent studies showed that at clinically relevant anaesthetic concentrations, MTs in neurons remain polymerized.

The mechanism of general anaesthesia is, we believe, an important clue to consciousness (e.g. Hameroff and Watt, 1983; Louria and Hameroff, 1995; Louria and Hameroff, 1996). Our view is that anaesthesia prevents quantum coherent superposition in hydrophobic pockets of a variety of neural proteins (receptors, ion channels, second messengers, enzymes, cytoskeletal MT subunits). Wulf and Featherstone (1967) showed that anaesthetic binding in internal protein hydrophobic pockets altered water binding at the protein outer surface, showing how the three types of quantum phenomena can be related.

G&C p. 24: ‘There is no evidence that quantum coherence involving super–radiance (or anything else for that matter) occurs in microtubules. At best, what Hameroff [and Penrose have] . . . done is to show that it might be possible. This should most definitely be distinguished from providing evidence that it is actual.’

P&H: Agreed. It is a model. Presumably, future technologies will either refute or verify it. To put this in perspective, we could parody G&C as follows: ‘There is no evidence that consciousness derives solely from activities at the level of neuronal assemblies. At best what the classical connectionists/reductionists have done is to show that it might be possible. This should most definitely be distinguished from providing evidence that it is actual.’

G&C p. 27: ‘Despite the rather breathtaking flimsiness of the consciousness–quantum connection, the idea has enjoyed a surprisingly warm reception, at least outside neuroscience. One cannot help groping about for some explanation for this odd fact.’ [After they isolate it, maybe they’ll find a vaccine against it. ‘Neuroscience’ can still be saved.] ‘Some people who, intellectually, are materialists, nevertheless . . . have a negative "gut" reaction to the idea that neurons . . . are the source of subjectivity and the "me-ness of me".’

P&H: Damasio (1994) suggests these people’s emotional sub-conscious, mediated from the brain through the autonomic nervous system in their ‘gut’, is trying to tell them something. Our model suggests the emotional sub-conscious may derive from ‘Platonic’ quantum computing (Shadows, p. 414). Perhaps the ‘explanation’ is that some people are capable of perceiving subconscious ‘gut’ feelings, and others are not!

G&C p. 27: ‘The crucial feature of neurons that makes them capable of processing and storing information is just ions passing back and forth across neuronal membranes through protein channels.’

P&H: Where is the evidence that neurons can process and store information without their cytoskeletons? More importantly, at this stage of the discussion, why can’t G&C say that these ion/membrane activities account for consciousness? It must be because, as they say earlier (p. 10), ‘Neuroscience has not reached the stage where we can satisfactorily answer these questions’. But even if every neuron, synapse, ion, channel and gene were mapped, would it necessarily tell us any more? Can the classical reductionists say whether C elegans (a nematode worm whose entire 302 neuron nervous system is completely mapped) is, or is not, conscious?

G&C p. 28: ‘Why should it be less scary, reductionist or counter–intuitive that "me–ness" emerges from the collapse of a wave function than from neuronal activity?’

P&H: Microtubule functions are neuronal activities. However, it is somewhat appealing to see how the phenomenon of consciousness could tie in with the behaviour of the universe at its deepest levels, and be relevant even to the very geometrical structure of spacetime.

G&C end by referring to Lewis Carroll’s Alice in Wonderland, claiming our model (at that time not yet published, nor reviewed by G&C): ‘is no better supported than any one of a gazillion caterpillar–with–hookah hypotheses.’

P&H:

‘It’s not that we’re in WonderlandBut p’raps their heads are in the sand.’

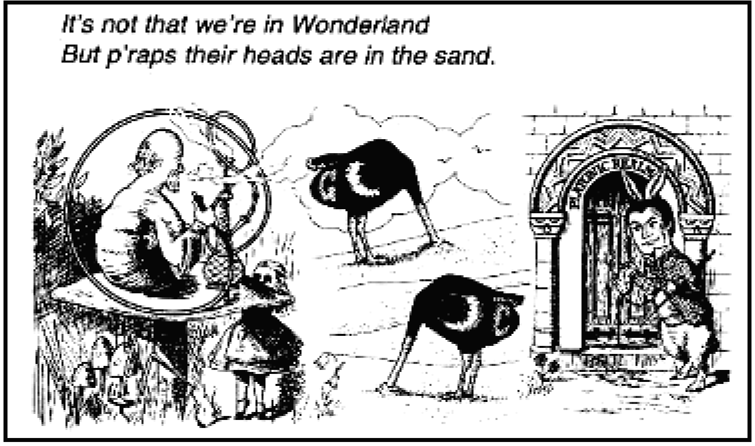
It's not that we're in Wonderland
But p'raps their heads are in the sand.

Figure 3: *By Jack Buckmaster, after Sir John Tenniel. Courtesy of Journal of Consciousness Studies.*

Conclusion

The Grush-Churchland arguments concerning the logical and physical parts of ‘Penrose’s viewpoint’ do not accurately express this viewpoint, nor do they at all take into account thorough discussion of such arguments already given in Shadows. On the biological side they exhibit serious misunderstandings, for example in relation to what they deem ‘physiological evidence’ regarding effects of the drug colchicine. Nonetheless, we thank them for directing attention to the Orch OR model, which we believe deals with the serious problems of consciousness more directly and completely than any previous theory.

References

Albrecht–Buehler, G. (1992), ‘Rudimentary forth of cellular "vision"’, Proc. Nat! Acad Sci. USA, 89 (17), pp. 8288-92.

Allison, A.C. and Nunn, J.F. (1968), ‘Effects of general anesthetics on microtubules: a possible mechanism of anaesthesia’, Lancet, ll, pp. 1326-9.

Anderson, M.II., Ensher, J.R., Matthews, M.R., Wiseman, C.E. and Cornell, E.A. (1995), ‘Observation of Bose–Einstein condensation in a dilute atomic vapor’, Science, 269, pp. 198 2Q1.

Aoki, C. and Siekevilz, P. (1985), ‘Ontogenic changes in the cyclic adenosine 3 , 51 monophosphate •stimulatable phosphorylation of cat visual cortex proteins, particularly of microtubule-associated protein 2 (MAP2): effects of normal and dark rearing and of the exposure to light’, J. Neurosci., 5, pp.2465-83.

Atema,1. (1973), ‘Microtubule theory of sensory transduction’, J. Theor. Biol., 38, pp. 181-90.

Beck, F. and Eccles, J.C. (1992), ‘Quantum aspects of brain activity and the role of consciousness’, Proc. Nail. Acad Sci. USA, 89 (23), pp. I 1357 1.

Bennett, E., Alberti, M.H. and Flood, l., (1981), ‘Uptake of [3H]Colchicine into brain and liver of mouse, rat, and chick’, Pharmacology, Biochemistry and Behavior, 14 (6), pp. 863-9.

Bensimon, G. and Chemat, R. (1991), ‘Microtubule disruption and cognitive defects: effect of colchicine on teaming behavior in rats’, Pharmacol. Biochem. Behavior, 38, pp. 141-5.

Bigot, D. and Hunt, S.P. (1990), ‘Effect of excitatory amino acids on microtubule-associated proteins in cultured cortical and spinal neurons’, Nervosci. Lett., 111, pp. 275-80.

Chou, K-C., Zhang C-T. and Maggiore, G.M: (1994), ‘Solitary wave dynamics as a mechanism for explaining the internal motion during microtubule growth’, Biapolymers, 34, pp. 143 53.

Clegg, J.S. (1983), ‘Intracellular water, metabolism and cell architecture’, in Coherent Excitations in Biological Systems, ed. H. Fröhlich and F. Kremer (Berlin: Springer-Verlag), pp. 162 75.

Damasio, A.R. (1994), Descartes’ Error: Emotion, Reason, and the Human Brain (New York: Putnam).

Dayhoff, I.E., Hameroft; S., Lahoz-BelUa, R. and Swenberg, C.E. (1994), ‘Cytoskeletal involvement in neuronal teaming: a review’, Eur. Biaphys. J, 23, pp. 79-93.

Del Giudice, E., Doglia, S. and Milani, M. (1983), ‘Self-focusing and ponderomotive forces of coherent electric waves: a mechanism for cytoskeleton formation and dynamics’, in Coherent Excitations in Biological Systems, ed. H. Fröhlich and F. Kremer (Berlin: Springer-Verlag), pp. 123-7.

Dustin, P. (1984), Microtubules (2nd Revised Ed., Berlin: Springer).

Eccles, J.C. (1992), ‘Evolution of consciousness’, Proc. Nail. Aced. Sci., 89, pp. 7320-d.

Emerich, D.F. and Welsh, T.1. (1991), ‘Ganglioside AGF2 prevents the cognitive impairments and cholinergic cell loss following intraventricular colchicine’, Experimental Neurology, 112, pp. 328-37.

Ergin, V. (1983), Magintnye Svoistva i Struktura Rasrvorov Elektritov (Moscow: Nauka, in Russian).

Friedrich, P. (1990), ‘Protein structure: the primary substrate for memory’, Neurosci., 35, pp. 1 7.

Fröhlich, H. (1968), ‘Long-range coherence and energy storage in biological systems’, Int. J. Quantum Chem., 2, pp. 641-9.

Fröhlich, H. (1970), ‘ Long range coherence and the actions of enzymes’, Nature, 228, p.1093.

Fröhlich, H. (1975), ‘The extraordinary dielectric properties of biological materials and the action of enzymes’, Proc. Nat! Acad Sci., 72, pp. 4211-15.

Grush R., Churchland P.S. (1995), ‘Gaps in Penrose’s toilings’, J. Consciousness Studies, 2 (1), pp. 10-29.

Halpain, S. and Greengard, P. (1990), ‘Activation of NMDA receptors induces rapid dephosphorylationof the cytoskeletal protein MAP-2’, Neuron,

Hameroff, S.R. and Walt, R.C. (1982), ‘Information processing in microtubules’, J. 77teor. Biol, 98, pp. 5491.

Hameroff, S.R. and Watt, R.C. (1983),’ Do anesthetics act by altering electron mobility?’, Anesth. Analg., 62, pp. 936-40.

Hameroff, S.R. (1987), Ultimate Computing: Biomolecular Consciousness and Nanotechnology (Amsterdam: Elsevier North-Holland).

Hameroff S. and Penrose R. (1996), ‘Orchestrated reduction of quantum coherence in brain microtubules - a model for consciousness’, in Toward a Science of Consciousness: Contributions from the 1994 Tucson Conference, ed: S Hameroff, A Kaszniak and A Scott (Cambridge, MA: MIT Press).

Hameroff S. and Penrose, R. (1995), ‘Orchestrated reduction of quantum coherence in brain microtubules’, Proceedings of the International Neural Network Society, Washington DC, July I7-11, 1995 (Hillsdale, N.J: Erlbaum).

Hameroff S. and Penrose, R (forthcoming) ‘Conscious events as orchestrated space-time selections’, Journal of Consciousness Studies.

Hirokawa, N. (1991), ‘Molecular architecture and dynamics of the neuronal cytoskeleton’, in The Neuronal Cytoskeleton, ed. RD Burgoyne (New York: Wiley-Liss), pp. 5-74.

Jibu, M., Hagan, S., Hameroff, S.R., Pribram, K.H. and Yasue, K. (1994), ‘Quantum optical coherence in cytoskeletal microtubules: implications for brain function’, BioSystems, 32, pp. 195 209.

Jibu, M., Yasue, K. and Hagan, S. (1995), ‘Water laser as cellular "vision"’, Submitted.

Kolasa, K., Jope, R.S., Baird, M.S. and Johnson, G.V.W. (1992),’Alterations of choline acetyltransferase, phosphoinositide hydrolysis, and cytoskeletal proteins in rat brain in response to colchicine administration’, Exp. Brain Res., 89, pp. 496-500.

Lee, J.C., Field, D.J., George, H.J. and Head, J. (1986), ‘Biochemical and chemical properties of tubulin subspecies’, Ann. NYAcad Sci., 466, pp. I 1 I 28.

Libet, B., Wright, E.W. Jr., Feinstein, B. and Pearl, D.K. (1979), ‘Subjective referral of the timing for a conscious sensory experience’, Brain, 102, pp. 193-224.

Louria D. and Hameroff, S. (1996), ‘Computer simulation of anesthetic binding in protein hydrophobic pockets’, in Toward o Science of Consciousness: Contributionsfrom the /994 Tucson Conference, ed. S.

Hameroff, A. Kaszniak and A. Scott (Cambridge, MA: MIT Press).

Matsuyama, S. S. and Jarvik, L.F. (1992), ‘Hypothesis: Microtubules, a key to Alzheimer’s disease’, Proc. Nat. Acad. Sci. USA, 86, pp. 8152.

Penrose, R. (1989), The Emperor’s New Mind (Oxford: Oxford University Press).

Penrose, R. (1994), Shadows ojthe Mind (Oxford: Oxford University Press).

Pribram, K.H. (1991), Brain and Perception (Hillsdale, N.1: Lawrence Erlbaum).

Puck, T.T. and Krystosek, A. (1992), ‘Role of the cytoskeleton in genome regulation and cancer’, Int. Rev. Cytology, 132, pp. 75-108.

Putnam, H. (1994), ‘The best of all possible brains?’, review of Shadows of the Mind, New York Times Review of Books, November.

Rasmussen, S., Karampurwala, H., Vaidyanath, R., Jensen, K.S. and Hameroff, S. (1990), ‘Computational conitectionism within neurons: a model of cytoskeletal automata subserving neural networks’, Physica D, 42, pp. 428-49.

Ricciardi, L.M. and Umezawa, H. (1967), ‘Brain and physics of many-body problems’, Kybernetic, 4, pp. 44-8.

Roth, L.E. and Pihlaja, D.J. (1977), ‘Gradionation: hypothesis for positioning and patterning’, J. Protozoology, 24 (1), pp. 2-9.

Sataric, M.V., Zakula, R.B. and Tuszynski, J.A. (1992), ‘A model of the energy transfer mechanisms in microtubules involving a single soliton’, Nanobiology, 1, 445-56.

Theurkauf, W.E. and Vallee, R.B. (1983), ‘Extensive cAMP-dependent and cAMP-independent phosphorylation of microtubule associated protein 2’, J. Biol Chem., 258, pp. 7883-6.

Tuszynski, J., Hameroff, S., Salaric, M.V., Trpisova, B. and Nip, M.L.A. (1995), 'Ferroelectric behavior in microtubule dipole lattices; implications for information processing, signaling and assembly/disassembly', J. Theor. Biol., in press.

Uedaira, H. and Osaka, A. (1989), Water in Biological Systems (Tokyo: Kodansha Scientific, in Japanese).

Vassilev, P., Kanazirska, M. and Tien, H.T. (1985), 'Intermembrane linkage mediated by tubulin', Biochem. Biophys. Res. Comm., 126, pp. 559-65.

Wang, N. and Ingber, D.E. (1994), 'Control of cytoskeletal mechanics by extracellular matrix, cell shape and mechanical tension', Biophysical Journal, 66 (6), pp. 21819.

Woolf, N.J., Young, S.L., Johnson, G.V.W. and Fanselow, M.S. (1994), 'Pavlovian conditioning alters cortical microlubule-associated protein-2', NewoReport, 5, pp. 1045-8.

Wulf, R.J. and Featherstone, R.M. (1957), 'A correlation of van der Waals constants with anesthetic potency', Anesthesiology, 18, pp. 97-105.

Yu, W. and Baas, P. W. (1994), 'Changes in microtubule number and length during axon differentiation', J Neuroscience, 14 (5), pp. 2818-29.

Acknowledgment
The authors are grateful to Vanessa Thomas for advice regarding the manuscript.Endnotes

1. It is extremely frustrating, considering the efforts involved in writing a book with particularly detailed arguments, when these arguments are simply treated as though they did not exist! Grush and Churchland are by no means unique in their ignoring most of the detailed arguments given in Shadows. For example, in their footnote 9, page 13, they refer to the 'powerful criticism of Putnam' (1994), who equally seems not even to have read the arguments in Shadows of relevance to his discussion. There is an unfortunate error in Shadows which arises from a misunderstanding on the part of the author about the precise relationship between the specific statement that Gödel originally produced and the condition of Ω -consistency for a formal system, for which the notation 'Ω (F)' is used. This has been corrected in the paperback edition, but the alteration makes virtually no change to the arguments. The simplest way for a reader in possession of the original hardback printing to deal with the point is simply to replace each occurrence of 'Ω (F) in Chapter 3 with G(F).
2. The faster emergence we describe is distinct from 'reflex' phenomena: touching your finger to a hot stove may result in spinally mediated withdrawal before conscious awareness of the pain.